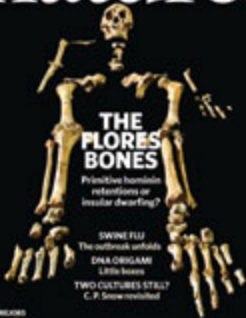


# nature



## THE FLORES BONES

Primitive hominid  
relictions or  
insular dwarfing?

SWINE FLU

The outbreak unfolds

DNA ORIGAMI

Little boxes

TWO CULTURES STILL?

C. P. Snow revisited

## Abstractions



### LAST AUTHOR

DNA molecules can be folded to create arbitrary two-dimensional shapes, such as smiley faces and stars (P. W. K. Rothemund *Nature* 440, 297–302; 2006). This

pioneering discovery by Paul Rothemund has inspired many, among them Jørgen Kjems, a molecular biologist at Aarhus University in Denmark. He and his colleagues set out to see whether they could translate Rothemund's 'DNA origami' into three dimensions (see page 73). Kjems tells *Nature* about prompting DNA molecules to self-assemble into a hollow, nanometre-scale box.

### How did you get DNA molecules to form a box shape?

All single-stranded DNA molecules can self-assemble, or direct themselves into a particular structural configuration. We first used a computer program to predict what sequences to synthesize that would direct the molecules to assemble into this structure. Then, we used one long, naturally occurring molecule from a virus, and about 220 short artificially synthesized molecules that bound to the long molecule to build a box 42 by 36 by 36 nanometres. This method can be used to assemble any shape — our DNA box is just the beginning.

### Can you put anything into the box?

It may be possible to put an enzyme into the box that produces a signal only when the lid is open and the substrate becomes available. The box could effectively become a sensor to signal the presence of a gene from a virus or a bacterium, for example. We are also experimenting with hiding a drug in the box that can kill a cell when the box opens. And we think that the box could potentially be used to make simple arithmetic calculations. Thus, if you have many boxes, you can make very complicated calculations — or effectively create a DNA computer.

### Were there surprises along the way?

What surprised me is that nature can direct self-assembly so nicely. I still don't understand how the process actually works. The biggest struggle we faced was working with the DNA itself. DNA isn't very stable because it is easily degraded by enzymes, which tear it into pieces. We're working on trying to use unnatural building blocks, by chemically altering the nucleotides that make up DNA, to make these structures more stable.

### Is this a popular field?

Yes. We just managed to be the first group to publish a complex three-dimensional DNA structure — there is a wave of similar experiments going on worldwide. This is intriguing enough that you'd say, 'let's try it.' ■

## MAKING THE PAPER

Hendrikje Nienborg & Bruce Cumming

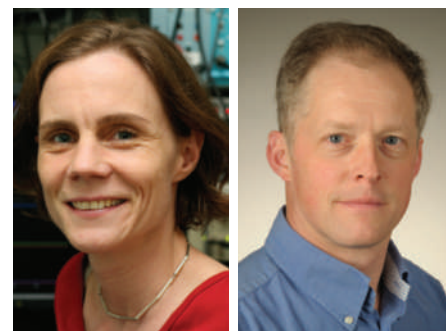
### The brain 'tampers' with incoming sensory data to fit expectations.

Our brains decipher a wealth of sights, sounds and other sensory information to allow us to make sense of our environment. But interpreting sensory inputs isn't always straightforward, as anyone who has stared at the infamous 'young girl–old woman' illusion knows. Hendrikje Nienborg and Bruce Cumming, working at the National Institutes of Health in Bethesda, Maryland, have now discovered that the brain 'tampers' with the signals it receives to favour one interpretation over another.

Signals from sensory organs activate sensory neurons, which in turn relay those signals to the brain areas that decode them and act upon the information provided. Researchers have known for more than a decade that the activity of sensory neurons varies not only in response to a particular stimulus, but also according to how the brain ultimately interprets that stimulus. "Imagine you are walking in thick fog, looking for a friend wearing a green leather jacket," says Nienborg, who was a postdoc in Cumming's lab before moving to the Salk Institute for Biological Studies in La Jolla, California. "When a blob appears in front of you, you have to decide whether or not the blob is a green jacket." The activity of the sensory neurons will vary depending on whether the decision is 'yes' or 'no'.

The widely accepted explanation for such variation is that sensory neurons have a role in decision-making. To test this idea, Nienborg recorded the electrical activity of individual sensory neurons in two monkeys as the animals performed a simple task. Each monkey was shown a series of dot patterns on a computer screen, and had to decide whether the centre of a circular pattern was protruding or receding.

The work, which provided a detailed description of the sensory neurons' activities during



Hendrikje Nienborg (left) and Bruce Cumming.

the course of each trial, was painstaking. "One downside of the technique is that it requires a lot of data, so we had to do many, many trials for each neuron," says Nienborg. Although individual trials took only a couple of seconds, the monkeys performed between 800 and 900 trials for each neuron studied. And the data generated were not always of sufficient quality. "You could find out at the end of a week of recording all day, every day that you did not have any data," says Cumming. "On the other hand, the next week you could have recordings from five neurons."

In the end, they were able to analyse data from 76 neurons. But the results were not consistent with the conclusion that sensory neurons have a direct effect on decision-making (see page 89). Instead, Nienborg and Cumming concluded that brain areas involved in decision-making are sending signals to sensory neurons, altering their activity. "What we have shown is that the brain changes the sensory input," says Cumming. "In a way, the brain is tampering with the data." To explain the findings in terms of the fog analogy, Nienborg adds: "What this means is that if we expect to see a green jacket we are more likely to see a green jacket."

From an evolutionary perspective, this might indicate that, when faced with uncertain sensory information, it is better to commit to one interpretation or another than to hesitate. If you aren't sure where a predator is coming from, "it might be better to make the decision to run left and get it wrong 50% of the time than to just stand there and get killed every time," says Cumming. ■

## FROM THE BLOGOSPHERE

When it comes to measuring the challenges of the 'human dimensions' of climate change, it seems social scientists will be taking centre stage. That was a key opening message from the International Human Dimensions Programme on Global Environmental Change Open Meeting in Bonn, Germany, 26–30 April, reports Anna Barnett on the blog Climate Feedback (<http://tinyurl.com/ceekvx>).

Barnett, assistant editor of *Nature Reports: Climate Change*, caught up with one of the keynote speakers, physicist Hans Joachim Schellnhuber, at a coffee break. He told her that physicists can describe climate threats increasingly vividly, but that it's up to social scientists to figure out how we bring about massive economic and social transformation.

For example, the technical problems with transferring solar power from the Sahara to Europe are already solved. It's the lack of legal frameworks, intergovernmental agreements and international will that stands in the way, Schellnhuber said. He urged social scientists to take the lead and to rethink their research scales from the local case study to globe-spanning projects. ■

Visit Nautilus for regular news relevant to *Nature* authors ▶ <http://blogs.nature.com/nautilus> and see Peer-to-Peer for news for peer reviewers and about peer review ▶ <http://blogs.nature.com/peer-to-peer>.

# Between a virus and a hard place

Complacency, not overreaction, is the greatest danger posed by the flu pandemic. That's a message scientists would do well to help get across.

**D**amned if you do, damned if you don't. The emergence of a new, swine-flu-related H1N1 strain of influenza in people in North America, with sporadic cases elsewhere in the world, has left the US Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia, and the World Health Organization (WHO) in Geneva in an unenviable position.

For more than a week now, these two agencies have been holding daily media briefings to keep the world informed about the rapidly unfolding story. There is ample reason for concern: a new flu virus has emerged to which humans have no immunity, and it is spreading from person to person. That has happened only three times in the past century. The pandemics of 1957 and 1968 were mild in most people but still killed many, and that of 1918 — which also seemed mild in its early phases — killed at least 70 million people worldwide. As *Nature* went to press, the WHO had already upped its pandemic threat level from 3 to 5, and a final step to its highest level of 6 — a global pandemic — seemed only a matter of time.

Yet at this early stage, the consequences of the pandemic are so uncertain that communicating the risks is a delicate matter. Influenza viruses evolve rapidly, making it extremely difficult to predict what this strain might look like a few months from now. If the agencies alert people and the pandemic fizzles out, they will be accused of hyping the threat and causing unnecessary disruption and angst. Indeed, just such a media backlash is already beginning, because most cases so far have been mild. But if the agencies downplay the threat and an unprepared world is hit by a catastrophe on the scale of 1918, the recriminations will come as fast as you can say 'Hurricane Katrina.'

To their credit, the WHO and the CDC have avoided the kind

of falsely reassuring officialese that has too often accompanied past crises. As Peter Sandman, a risk-communication consultant based in Princeton, New Jersey, aptly puts it: "Anyone who's paying attention gets it that we just don't know if this thing is going to fizzle, hang in abeyance for months, disappear and then reappear, spread but stay mild, replicate or exceed the 1918 catastrophe, or what. The reiteration of uncertainty and the insistence on what that means — e.g., advice may change; local strategies may differ; inconsistencies may be common — has been almost unprecedentedly good."

**"The risk is not hyping the pandemic threat, but underplaying it."**

Also encouraging is that many governments now have at least some kind of pandemic plan in place, thanks to the scare over the H5N1 avian flu virus earlier this decade. Five years ago very few of them did. But many of those plans contain an important element that has been conspicuously absent in the current communication by governments and public-health authorities: during a severe pandemic, there is only so much they can do. Much of the response will depend on local communities taking action for themselves.

Scientists can help, by serving as credible voices to inform their communities of the risks and uncertainties, and by pointing people to the pandemic-planning resources on the CDC and WHO websites, the PandemicFlu.gov site, and many others. For the moment, the risk is not hyping the pandemic threat, but underplaying it. We know a tsunami is coming. No one can say whether it will be just a large wave, or a monstrous one, but it is time to start thinking about at least being ready to move to higher ground. ■

## Keep to the vision

The United States should not try to keep its space shuttles flying beyond 2010.

**P**erhaps the most memorable achievement of the US space shuttle fleet was *Endeavour's* first servicing mission to the Hubble Space Telescope in December 1993. Here was something that only humans in space could achieve — the gentle snagging of the satellite with the shuttle's arm; the five spacewalks to install corrective optics, new instruments, gyrosopes and solar panels; and finally, the release of a clear-eyed telescope capable of discerning finer details, deeper in the cosmos, than any other telescope in history.

On 11 May (see page 21), the space shuttle *Atlantis* is scheduled to make what is not just the last trip to the Hubble, but also the last space-shuttle trip to any destination other than the still-rather-

pointless International Space Station, which is due seven more visits before shuttle flight operations cease in September 2010. As such, the Hubble flight marks — or should mark — the end of an era. Unfortunately, Congress has started making moves to keep the shuttle flying into 2011. The administration of President Barack Obama should resist this idea — and at the same time, take the opportunity to state more clearly its objectives for NASA and human spaceflight.

The "vision for space exploration" articulated by former US president George W. Bush in 2004 called for the shuttle to be replaced by an all-new system, now called Constellation, in which human crews would ride into space inside a bell-shaped Orion capsule mounted atop an Ares rocket. Constellation is in some ways a reconceived version of the Apollo infrastructure, and as such can seem technologically retrograde. But unlike the shuttle, it opens up the possibility of missions beyond low-Earth orbit — missions to the Moon, nearby asteroids and perhaps Mars.

If a nation must have human spaceflight — and it seems that, in terms of practical politics, the United States must — then travel to such interesting places seems the best of the available goals. That makes Constellation the way forwards. There is, however, a significant gap between the last shuttle flight next year and the first launch of an Orion capsule to the space station, which might not take place until 2015. Such a gap was planned; indeed, it was one of the space vision's braver aspects, recognizing that creating a new system while running the old one was not a viable option. Yet delays in Constellation's development mean that the gap is getting longer, leaving the United States dependent on Russia to deliver people to the space station in the interval. That perceived ignominy, combined with the job upheaval threatened by the demise of the shuttle, seems to some in Congress a reason to stretch out the shuttle's old age.

It is not. The basic logic remains sound: closing one human spaceflight operation before the next is functional saves money and ensures focus. NASA's next administrator, whenever he or she is finally named and confirmed, should be free to concentrate on meeting the goals that Constellation was designed for — without the distraction of also having to drag out the shuttle programme.

That administrator should also make it clear that human spaceflight, although a huge part of NASA's legacy, is not the overarching justification for the agency's existence. NASA also has a great deal of urgent work to do in studying Earth from orbit — and in providing new visions of the Universe beyond Earth. Servicing the Hubble might have been the best thing the shuttles did. But the images and knowledge provided by the Hubble, more than the mostly routine missions of the shuttle, are closer to the essence of what the space programme is there for. ■

## Doing good, 50 years on

Its attack on poverty and arrogance is what makes C. P. Snow's 'two cultures' lecture relevant today.

**I** did not expect much. Plenty of people were saying similar things. It seemed to me to be a time when one should add one's voice." Thus did the English scientist-turned-administrator-turned-novelist Charles Percy Snow recall his anticipations 50 years ago today, when giving his 1959 Rede lecture on 'the two cultures'. But by the time of his 1964 essay 'The Two Cultures: and a Second Look', it was clear that a nerve had been struck in many countries.

Many aspects of his lecture refer to a time long gone. Its attack on the lack of intellectual interest in the world of applied science and trade can hardly be said to apply now in his home country, transformed as it has been in its attention to technology transfer and investment. Michael Frayn, Tom Stoppard, John Updike and others have shown, in vivid and diverse contrast to his era, how the literary imagination has been fired by science (see page 34).

Many senior policy-makers might still have difficulty in reciting the second law of thermodynamics — Snow's notorious test of essential and neglected scientific awareness. But, thanks to extensive public discussion and access via the Internet, a great many of them are well aware of what research has to say about issues that immediately concern them, in climate, stem cells and much else.

Indeed, the boundaries between the arts and the sciences — and between the sciences themselves — are more porous than ever. Yes, multidisciplinary poses some substantial challenges in some domains (see page 32), especially when it comes to training. But those who bring disciplines together in the pursuit of scientific and technical answers find that the best people to have on board are those with deep specialized knowledge, and that such individuals can usually find a place in well-led collaborations.

One seldom-discussed aspect of Snow's talk is his discussion of the rich and the poor — a subject he subsequently wished he had expressed better. "With good fortune," he wrote in his 1964 essay, "we can educate a large proportion of our better minds so that they

are not ignorant of ... the remediable suffering of most of their fellow humans, and of the responsibilities which, once they are seen, cannot be denied."

It is in this territory alone that Snow's lecture truly has an edge today. On the one hand he would have celebrated the descendants of the physicists such as Ernest Rutherford whom he insightfully wrote about elsewhere — the purest of the pure in their scientific goals. The combination at CERN — Europe's particle-physics laboratory near Geneva, Switzerland — of world-class engineering with ambitions to seek deeper levels of the laws of nature would have inspired Snow.

Yet Snow would not have approved of the narrow-mindedness of some researchers who consider the significant costs of their work to be no more than their due from society, nor of their blind resentment when its value is questioned. *Nature* champions expensive science that seeks to answer the deepest questions about matter and the Universe. But it also endorses a spirit, as expressed in the final chapter of 'The Two Cultures', that gives just as much priority to scientists' responsibility to address humankind's most pressing problems.

What Snow urged in particular was an awareness of the problems of poor countries — and of putting scientists at the disposal of solving those problems, for reasons both moral and strategic. He expressed the disparities between rich and poor countries in terms of average lifespan. Those disparities may have shrunk but are still unacceptably large. Today's global threats may make them worse.

Snow's overriding messages — whether about awareness of artistic and scientific experience or about the applied sciences or about 'remediable suffering' — was that the best and the brightest should not be blinkered. That message still has resonance. Narrow-mindedness and any intellectual arrogance that lies behind it remain as unforgivable now as they were half a century ago. ■

**Join the debate at Nature Network at [www.tiny.cc/opinion191](http://www.tiny.cc/opinion191) to add your views and comment on these articles.**

**"Nature champions expensive science that seeks to answer the deepest questions about matter and the Universe."**



## RESEARCH HIGHLIGHTS

**Now hear this, or not**

PLoS One doi: 10.1371/journal.pone.0005413 (2009)

Although many mammals can hear very high frequencies, other vertebrates are less au fait with ultrasound.

Victoria Arch, of the University of California, Los Angeles, and her colleagues now report that a frog from Borneo, *Huia cavitympanum*, is the first non-mammalian vertebrate discovered to communicate with calls purely in the range above 20 kilohertz, which is about the upper limit of human hearing. This species had been known to produce these ultrasonic calls, and when playing them back in the field, the team found that male frogs nearby increased the frequency of their calls in response.

On examining the frog's brain and ears, the researchers showed that its hearing was most sensitive above 20 kilohertz.



T. ULMAR GRAFE

**NEUROGENETICS****Psychosis genes exposed**

Science 324, 605 (2009)

Scientists in Germany have provided the first evidence that genetic risk for psychotic illnesses is linked, at least in part, to abnormal connections between different brain areas.

Andreas Meyer-Lindenberg of the University of Heidelberg and his colleagues studied 115 healthy people with or without a particular variant of the gene *ZNF804A*. The variant has been identified in genome-wide association studies as possibly conferring a small risk for developing schizophrenia or bipolar disorder. The authors performed brain imaging while volunteers carried out cognitive tasks relevant to these disorders.

The team found that although the risk gene does not influence the strength of activation in various brain areas, connectivity between some areas is either reduced or increased in risk-gene carriers — in a pattern reminiscent of that seen in patients.

**VIROLOGY****HIV at the gates**

Cell 137, 433–444 (2009)

HIV enters human cells through a more complex pathway than previously predicted.

Researchers had long thought that HIV binds to cell-surface receptors, and then fuses directly with the cell membrane, dumping its dangerous payload into the cell within about 10 minutes.

Gregory Melikyan and his colleagues at the University of Maryland School of Medicine in Baltimore show that successful infection includes an additional step, in which the virus becomes enveloped by membrane and

internalized by the cell through a process called endocytosis. Thirty minutes to an hour can pass before the virus fuses with the internalized membrane and delivers its genetic material.

The findings may necessitate re-evaluation of drug candidates meant to block HIV's entry into cells.

For a longer story on this research, see <http://tinyurl.com/d5nuze>

**DNA REPAIR****Chemo's modus operandi**

PLoS Biol. 7, e1000091 (2009)

The chemotherapeutic 5-fluorouracil has been a first-line treatment for diseases such as colorectal cancer for decades, even though it's not clear exactly how the drug works.

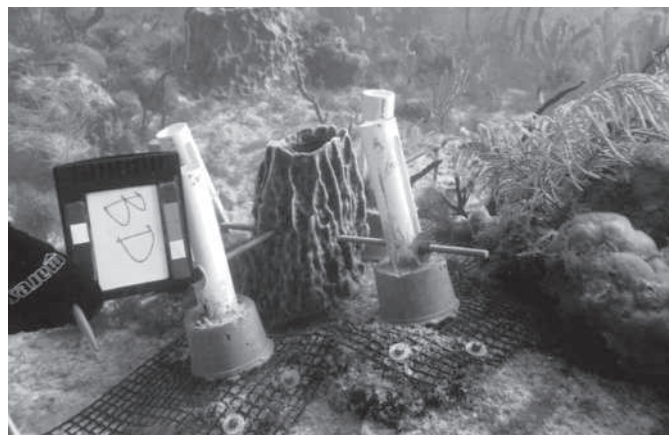
In cells, 5-fluorouracil is converted into several metabolites that mimic the natural RNA base uracil. The metabolites inundate the cell, some becoming incorporated into RNA and even DNA in dividing cells. Uracil in DNA activates a family of DNA-repair proteins devoted to removing it, but the unnatural 5-fluorouracil triggers incomplete repair, causing an accumulation of strand breaks in the DNA and eventual cell death. Primo Schär of the University of Basel in Switzerland and his colleagues examined mouse and human cells responding to the drug and identified thymidine DNA glycosylase (TDG) as the uracil-removing protein responsible for the breaks. Without TDG activity, cancer cells become more resistant to the drug.

**CONSERVATION****Reef repair**

Restor. Ecol. 17, 192–195 (2009)

Some large marine sponges, such as the barrel sponge *Xestospongia muta*, can live for hundreds of years. But when dislodged from the reefs they inhabit by storms, ship groundings or fishing lines, these organisms have little chance of reattaching naturally. Steven McMurray and Joseph Pawlik of the University of North Carolina in Wilmington have stumbled across a new method of reattaching them that could be useful in conservation efforts.

The duo skewered sponges with two perpendicular steel rods, then secured the rods to bases made from PVC piping, concrete and mesh that had been nailed to the reef's limestone bed (pictured below). The method, designed for temporary experiments, surprised the researchers by helping half of the 40 transplanted sponges to reattach, despite three passing hurricanes. The apparatus was removed once attachment was complete.



S. MCMURRAY, UNIV. NORTH CAROLINA, WILMINGTON

## MATERIALS SCIENCE

## Conductors with a twist

*Nature Mater.* **8**, 421–426 (2009)

Discotic liquid crystals are flat-cored organic molecules that stack in twisting, electron-conducting columns. They are used in photovoltaics and field-effect transistors. The best discotic species currently in use make stacks with a twist angle of 30°, but calculations show 60° to be optimal for conduction.

Klaus Müllen and Denis Andrienko of the Max Planck Institute for Polymer Research in Mainz, Germany, and their colleagues have synthesized a new molecule with a 60° twist. This doubled electron mobility, and through molecular dynamics simulations, the authors show that removing defects in the stacking structure could push that value higher.

## COSMOLOGY

## No ring or reason

*Astrophys. J.* **696**, 694–700 (2009)

A new study raises questions about the ring of dark matter reported to exist inside a galaxy cluster.

The cluster, called Cl0024+17, is actually thought to be two merging clusters, and some suggest that the ring was caused by their collision. To test the idea, John ZuHone, now at the Harvard-Smithsonian Center for Astrophysics in Cambridge, Massachusetts, and his colleagues simulated cluster collisions under varying conditions. They could not produce a ring unless the dark-matter particles had circular orbits.

Because such orbits are unlikely to occur in galaxy clusters, the results suggest that the ring report is questionable, says ZuHone.

## PLANT PHYSIOLOGY

## Gifts from grafts

*Science* **324**, 649–651 (2009)

Plants grafted together exchange genetic information, suggesting a new and surprising mechanism for gene transfer between organisms.

Grafting is commonly used in cultivation and can occur naturally when shoots or roots from different trees come into contact, but grafting was not thought to involve any mixing of genetic material. To test this, Sandra Stegemann and Ralph Bock of the Max Planck Institute for Molecular Plant Physiology in Potsdam-Golm, Germany, grafted together two transgenic tobacco plants expressing different antibiotic-resistance genes.

The resistance genes were frequently exchanged between cells across the graft site. However, transfer only occurred when the

genes were carried in the chloroplast genome, not when a resistance gene was inserted into the nuclear genome. Because the genetic exchange was limited to the graft site, the genes would only be passed to offspring of shoots formed at that site.

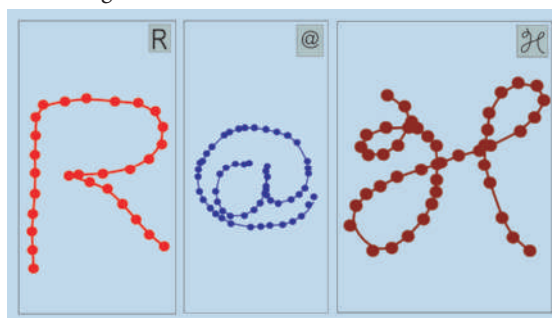
## STRUCTURAL BIOLOGY

## A virus laid bare

*PLoS Biol.* **7**, e1000092 (2009)

Mimivirus was once mistaken for a bacterium owing to its shape and great size. Michael Rossmann of Purdue University in West Lafayette, Indiana, and his colleagues have now revealed the structure of this largest of known viruses. Enzymatic scrubbing of long fibres that extend from the viral capsid gave the researchers a new view. They reconstructed the icosahedral shape of the virus using cryogenic electron microscopy and atomic force microscopy.

Mimivirus is not as symmetrical as many other large, double-stranded DNA viruses, in part because of a 'starfish' feature that appears on one side. This feature may allow the virus to deliver its genome into host cells.



## NANOTECHNOLOGY

## The helix that delivers

*Nano Lett.* **10**, 1021/nl900186w (2009)

For nanomedical applications such as delivering drugs or performing microsurgical procedures at specific places in the body, the ability to navigate bodily fluids with a high degree of control is crucial.

With this in mind, Ambarish Ghosh and Peer Fischer of the Rowland Institute at Harvard produced screw-like glass structures 1–2 micrometres long by nanofabrication. Depositing a thin layer of a ferromagnetic material on one side of each helix allowed the authors to direct the screws' motion with rotating magnetic fields. The duo even used the swimmers to spell out the initials of their institution (pictured above).

The microswimmers can carry chemicals and push loads, and could be used as probes in rheological measurements difficult to perform by other techniques.

## JOURNAL CLUB

David Kirchman

University of Delaware, Lewes

## A microbial ecologist learns something new from an old-fashioned study.

What could be easier than learning about an organism simply by watching how it varies over time in its natural habitat? You'd think this would have been done long ago for marine bacteria, which are important in many biogeochemical processes, including the carbon cycle; in fact, they're the organisms running the biosphere. But it's not easy to follow microbes in the open ocean, far from the lab and beyond the reach of standard techniques.

Craig Carlson at the University of California, Santa Barbara, and his colleagues took on this challenge for the most abundant group of marine bacteria: SAR11. They examined variations in SAR11 over several years in the Sargasso Sea, where the group was first discovered nearly 20 years ago (C. A. Carlson *et al.* *ISME J.* **3**, 283–295; 2009). Sequencing and other data had previously revealed that SAR11 bacteria are diverse and can account for almost 50% of microbes in a given marine environment; however, we still knew little about their natural history.

So Carlson's group looked to address a basic question: how do different members of SAR11 vary with depth and over time? They examined 13 years' worth of DNA samples, viewed 3 years' worth of preserved cells under the microscope, and then analysed the microbial data in light of what is known about SAR11's environment. Three SAR11 'ecotypes', they say, flourish differently at various depths and over a yearly cycle, which starts in spring, when deep mixing stops and photosynthesis speeds up.

The authors make good use of new genomic data from a lab-grown representative of SAR11 (*Pelagibacter ubique*) to understand Sargasso Sea populations, but the study's insight comes from the old approach of patiently watching organisms over time in their natural habitat.

Discuss this paper at <http://blogs.nature.com/nature/journalclub>



## SPECIAL REPORT

# How severe will the flu outbreak be?

Epidemiologists race to pin numbers on the global H1N1 spread.

The World Health Organization (WHO) this week remained on the verge of declaring a pandemic of the H1N1 swine-associated flu virus. Public-health bodies and scientists have made progress in starting to understand the outbreak, but major questions remain about how severe the disease will get.

As *Nature* went to press, the virus had extended its range to 21 countries on 5 continents, with 1,124 confirmed cases and 26 deaths. The WHO remained at pandemic alert phase 5 on a scale of 6, having moved up from 3 and then 4 following sustained human-to-human transmission in the Americas.

"We're well past the containment phase," says Ira Longini, an epidemiologist at the University of Washington School of Public Health in Seattle. "We are now in the mitigation phase."

The main challenge for scientists now is assessing how severe any pandemic might be. That's because the WHO's pandemic phases refer only to geographical spread of a new

disease; government pandemic plans look more to the outbreak's severity to guide their response. The United States, for example, has very different response plans depending on where a pandemic falls on its hurricane-like scale of severity, ranging from Category 1 (denoting case-fatality rates of less than 0.1%) to 5 (denoting case-fatality rates of 2% or above).

Most cases detected so far have been mild, with relatively few deaths for the number of cases observed. But reliable data are still lacking in Mexico, where most cases have occurred, and the number of cases in other countries remains too small statistically to detect levels of even 1–2% mortality rates, lower than the 1918 H1N1 influenza pandemic.

Researchers are, however, starting to work up early numbers for some other key variables of spread, including the basic reproductive rate,  $R_0$ , which is the number of new cases that an infected individual will give rise to. Longini's rough preliminary estimates from field data



Mobile health units check people with symptoms.

such as a school outbreak in New York, for instance, suggest that the current  $R_0$  is about 1.4. "At this point the virus does not seem as transmissible as past pandemic strains," he says. The  $R_0$  of the 1918 pandemic has been estimated at less than four; that of seasonal influenza typically ranges from 1.5 to 3.

Longini also suggests that the generation time — the time period before an infected individual begins infecting others — is probably between 3 and 5 days, but closer to 3. The higher the  $R_0$ , and the shorter the generation time, the faster and more difficult the spread is to control.

A related way of looking at the virus is the secondary-attack rate, which can be estimated from field data by looking at what proportion of a relatively confined cohort such as a household or school falls ill after exposure to a contact. The current secondary-attack rate is around 25–30%, according to Anne Schuchat of the US Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia. But this number, too, could change.

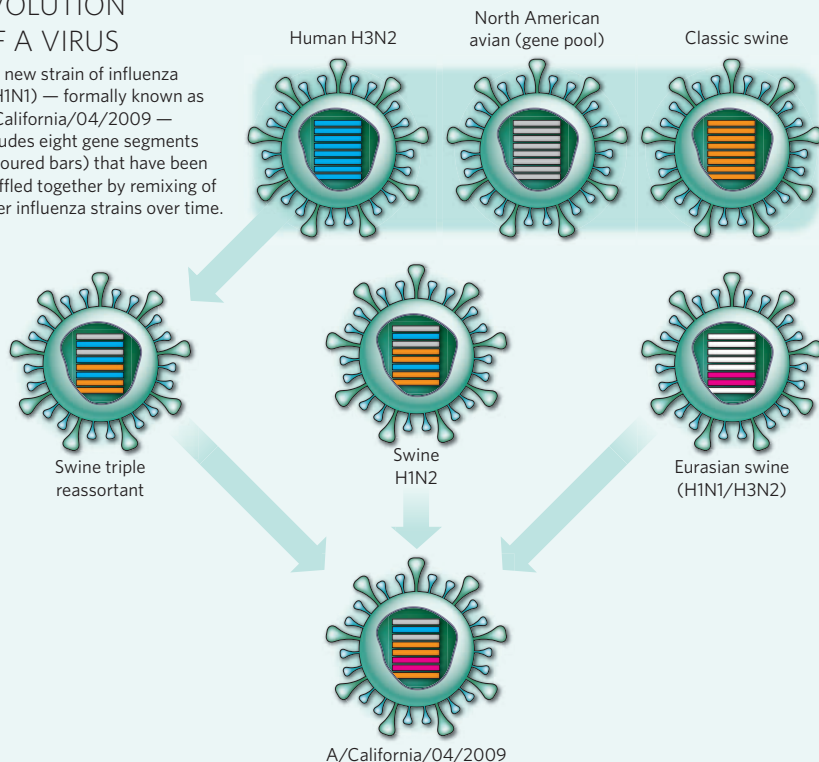
Just two weeks into the outbreak, researchers are using the incoming data on transmissibility and infectivity to model scenarios of how the virus might evolve, and to refine and target mitigation-control measures for varying levels of virulence, says epidemiologist Marc Lipsitch of the Harvard School of Public Health.

"My own personal view is that it is too early to say what the severity is," says Lipsitch, who works with the CDC as part of a 'Team B' made up of academic researchers and other outside advisers. One reason the disease appears mild is that in the United States there has been only one

SOURCE: G. SMITH

## EVOLUTION OF A VIRUS

The new strain of influenza A(H1N1) — formally known as A/California/04/2009 — includes eight gene segments (coloured bars) that have been shuffled together by remixing of other influenza strains over time.





O. TORRES/AFP/GETTY IMAGES

Morse, an epidemiologist at Columbia University in New York City. “Increased virulence — seeing a number of more serious cases — especially in healthy young adults, would be a trigger for a greatly increased response.” As of 4 May, 62% of cases in the United States were in people under the age of 18.

In Mexico, authorities said on 4 May that they would lower the alert level and begin lifting some restrictions on public gatherings, including the shuttering of restaurants. Longini describes the country’s actions as appropriate to stemming the spread of a new disease. “Mexico has done a great job of taking the necessary social distancing and containment methods needed to help dampen transmission,” he says. Mitigation measures are also intended to minimize social disruption and maximize availability of medical care — by trying to avoid everyone being off sick from work, or overwhelming hospitals, at the same time.

Crucially, they also can buy time until vaccine production can be ramped up. The WHO is looking at how a vaccine for the currently circulating H1N1 strain could be produced, including possibly curtailing the production run for the upcoming seasonal influenza vaccine in order to ramp up production on a vaccine for the new strain. The earliest an H1N1 vaccine could be made would be in four to six months.

For now, the disease is treatable with the drugs zanamivir (Relenza) and oseltamivir (Tamiflu).

**Declan Butler**

reported death among 286 cases. But such case-fatality rates, reported one week after a case is confirmed, can underestimate actual mortality rates, because they can overlook patients who remain ill and subsequently die. This was demonstrated in 2003, when the reported case-fatality rate of severe acute respiratory syndrome (SARS) in Asia was initially put at a few per cent — only to be found to be closer to 20% in studies that followed up cases.

Experts also remember that the 1918 pan-

demic began with a wave of mild disease in early 1918, only to return with a vengeance at the end of the year. “I’m particularly concerned about what will happen in the fall,” Schuchat told reporters on 3 May.

Continually estimating mortality rates is now the top priority of public-health bodies and scientists, and the task should become easier as the number of cases grows.

“The uncertainty is a major challenge. We don’t predict influenza very well,” says Stephen

## The turbulent history of the A(H1N1) virus

Genetic analyses of the A(H1N1) strain of swine influenza indicate it is a mixture of at least two flu viruses that circulated separately through pigs in North America and Eurasia for 10–20 years, before mixing and jumping to humans.

Flu viruses mix in pigs all the time, and the event that gave rise to this particular strain probably occurred before the beginning of 2009 — most likely around September 2008 — estimate an international team of researchers led by Andrew Rambaut of the University of Edinburgh, UK, Oliver Pybus of the University of Oxford, UK, Michael Worobey of the University of Arizona in Tucson and Gavin Smith of the University of Hong Kong (see <http://influenza.bio.ed.ac.uk>). “This result would seem to concur with a low

virulence for the strain so far, since it means that the virus circulated for several months before being detected,” Pybus says.

These and other researchers, including Raul Rabadan of Columbia University in New York and Steven Salzberg of the University of Maryland, College Park, have found that six of the swine flu’s genetic segments probably come from a North American strain of swine flu. The other two segments seem to come from Eurasian swine flu strains (see graphic, left). The North American and Eurasian ‘parental’ strains contain genes from flu viruses that had previously infected birds and people.

Nancy Cox, director of the influenza division at the US Centers for Disease Control and

Prevention in Atlanta, Georgia, says that the virus seems to be much the same in different patients. “All of the genes of all of the viruses that we have examined to date are 99 to 100% identical,” Cox told reporters on 1 May. “It will be somewhat easier to produce a vaccine, because the viruses that are spreading are so similar to each other.”

Cox adds that the virus does not contain any of the genetic markers for virulence that have been seen in analyses of the pandemic 1918 flu strain. But, she notes, “there’s a great deal that we do not yet understand about the virulence of the 1918 virus or other influenza viruses that have a more severe clinical picture in humans.”

The fact that swine flu’s parental strains were predominantly found

in North America and Eurasia does not mean that the new virus originated in either place, because viruses and pigs often move across international borders.

Scientists will now watch to see whether the H1N1 strain mixes with seasonal human flu viruses. If people become infected with both swine and seasonal flu simultaneously, the two viruses could recombine into new ones with unpredictable properties. “There’s going to be some interesting dynamics ahead when [A(H1N1)] interacts with seasonal flu,” Rambaut says.

Others have even raised the possibility of the new strain reassorting with H5N1, the avian influenza virus that has been circulating since 2003.

**Erika Check Hayden**



# China joins world-class synchrotron club

Nation's costliest science facility is unveiled.

## SHANGHAI

The Shanghai Synchrotron Radiation Facility (SSRF) officially opened its doors last week to a queue of scientists waiting hungrily for beamline time. The 1.2-billion renminbi (US\$176-million) light source is China's biggest investment in a single science facility to date, says Zhao Zhentang, an accelerator physicist and the facility's deputy director.

The synchrotron radiation is generated as magnets bend an electron beam around the main ring's 432-metre circumference. "It's like mud coming off a spinning tyre," says Herman Winick, assistant director emeritus of the Stanford Synchrotron Radiation Laboratory in California, and chair of the committee of 30 international and 4 Chinese scientists who last week gave the go-ahead to approve the facility's first seven beamlines.

The light ranges from high-energy, hard X-rays (10 kiloelectronvolts and above)

through to infrared frequencies. It is channelled into a number of beamlines jutting out from the main ring, where it can be used for spectroscopy and diffraction experiments in fields including condensed-matter physics, structural biology and medical imaging.

In 2001, the Chinese government rejected plans for a new synchrotron light source, unconvinced that it would have enough users. Zhao and his colleagues resubmitted the proposal in June 2004 with the endorsement of scientists from more than 100 universities and some 20 institutes of the Chinese Academy of Sciences.

Since then, the project has progressed apace. Construction began on 25 December 2004, and the first synchrotron light was seen within three years, on 24 December 2007. "It's got to be a world record," says Winick. The facility is expected to have 30 working beamlines within 5 years, with the potential for up to 60 lines.

The Shanghai synchrotron is drawing Chinese scientists back from abroad.

More than 60 synchrotron radiation facilities operate worldwide, with a fresh crop of cutting-edge light sources in the works (see 'Leading lights'), giving fairly routine access to scientists in most countries that invest in such research.

Before the SSRF started up, China had two light sources, in Beijing and Hefei. But these are relatively small and cannot generate hard



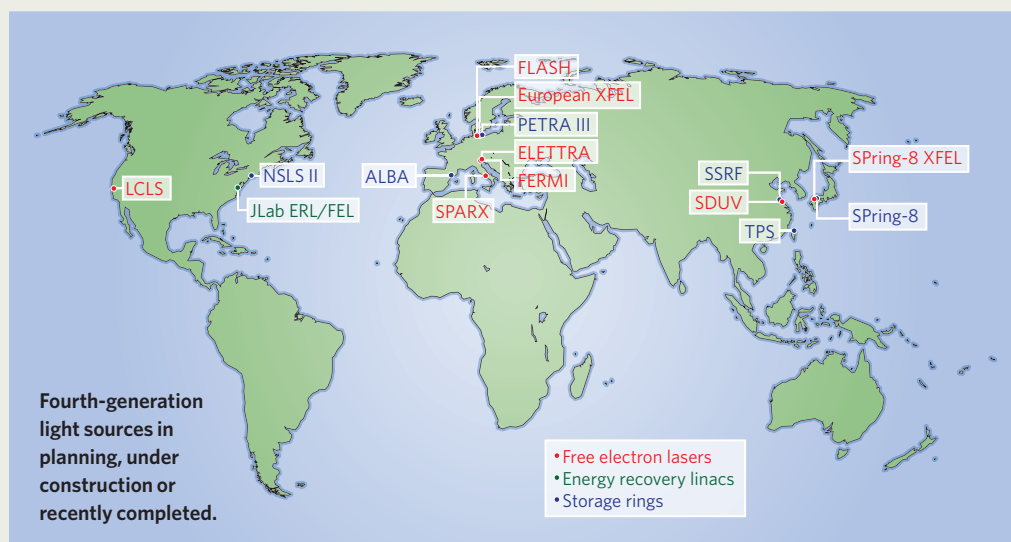
SHANGHAI INST. APPL. PHYS.

## Leading lights

China is not the only country brightening up its X-ray sources. Two other new light-source facilities, the Linac Coherent Light Source (LCLS) in Stanford, California, and the Positron-Electron Tandem Ring Accelerator (PETRA) III, in Hamburg, Germany, have opened in the past month. And more than a dozen are under way elsewhere (see map).

This 'fourth generation' of new facilities improves on older synchrotrons by making photon beams that are many orders of magnitude brighter, and delivering much shorter-duration light pulses.

Some of the new-generation facilities, called ultra storage rings, build on existing synchrotron technology. But others use approaches based on linear accelerators: free electron lasers (FEL) and energy recovery linacs (ERL). With bright, precisely timed pulses of light, they will allow scientists to create 'movies'



of atomic processes. They could reveal, for example, the step-by-step process of photosynthesis in a leaf, or the precise shape of a catalyst at the moment it becomes most chemically active.

Previous-generation light sources were something of a status symbol for countries

wanting to join the high-tech club, says Franz Himpsel, a physicist at the University of Wisconsin, Madison. But now countries are realizing that the research can translate into products that have an economic impact, says Steve Dierker, project director for the National Synchrotron

Light Source-II (NSLS-II) at Brookhaven National Laboratory in New York, which is due for completion in 2015. "Countries are investing in the basic scientific infrastructure to enable them to have a competitive edge in the marketplace," he says.

**Eric Hand**



X-rays. For many Chinese scientists, who have lacked either the funding or the visas to use facilities abroad, the SSRF is a chance to see what cutting-edge synchrotrons can do.

Zhao hopes that the investment in world-class facilities such as the SSRF will give Chinese scientists more reason to keep their research programmes in China, or to return from abroad. Indeed, Guo-Yuan Yang, a neurosurgeon who studied the mechanisms of cerebrovascular diseases for 20 years in the United States, says that he has returned to China in part because of the SSRF. Now based at Shanghai Jiao Tong University, he says the beams will allow him to examine, for example, real-time changes in small blood vessels seen in arteriovenous malformation — a circulatory disorder. “In the past we had to sacrifice animals, but here we can follow them day by day,” he says.

At another SSRF beamline dedicated to X-ray absorption fine structure (XAFS), which is particularly suited to analysing the atomic structure of liquids, a group led by Guozhong Wu of the Shanghai Institute of Applied Physics is looking at the iron ion content in an ionic liquid to investigate its catalytic properties. The institute's Yuying Huang, who is in charge of the beamline, says 60 research groups from China have already requested time. “We are asking the government to build another [XAFS] line,” he says.

Lili Chen of the Shanghai Institute of Materia Medica, who is using a beamline to look at the structures of complex proteins related to diabetes and other diseases, says there are already applications to use 7,000 hours of time on that beamline in 2010 — almost twice as much as has been budgeted for.

With demand that high, says Winick, “it's going to be a zoo here in a year”.

David Cyranoski

## Even big societies feel the pinch

The American Chemical Society (ACS), the world's biggest scientific society, is feeling the effects of the global economic downturn.

On 28 April, six months after tightening its belt a first notch, the society laid off 56 people, 3% of its employees. New employment will be frozen indefinitely, and the society has capped its contributions to medical insurance for retired employees and cut back on other expenses such as travel.

The ACS, based in Washington DC, blames its financial problems on the falling values of its investments, as well as falling revenue from its publications, which include the *Journal of the American Chemical Society* and *Chemical & Engineering News*, as subscribers move from print to online-only deals. Its reserves were US\$60 million at the end of 2008, compared with \$212 million the year before.

The society has also posted a \$36.5-million bond in a lawsuit it recently lost against the chemical-information company Leadscope in Columbus, Ohio. It had argued that the founders of Leadscope — former employees of the society — had used intellectual property belonging to the society. The ACS plans to appeal against the decision, but if it loses, it may owe that \$36.5 million plus interest. According to Glen Ruskin, director of the ACS office of public affairs, the possible outlay is “not having any impact on our financials at this point, nor did it figure in the actions we took on Tuesday [28 April]”.

Of the 56 layoffs, 40 were in the publications division. Ten staff members at *Chemical & Engineering News* and the entire reporting team at *Environmental Science & Technology*

were laid off. The publications division generates more revenue than any other division of the society apart from the Chemical Abstracts Service, a fee-based database of chemical information and suite of tools to search it.

According to notes taken by a laid-off employee at a meeting of the publications division on 30 January, the ACS earns about \$500 million annually in revenues, with 29% from the publications division. In 2008, revenue growth in the division was just 2.5%, down from 7–10% a year.

Ruskin says that the layoffs were strategically chosen. “The whole publishing enterprise has entered probably one of the more momentous times in its history: we

**“The whole publishing enterprise has entered a momentous time.”**

are now moving from print to electronic,” he says. “Part of what we are doing with the reduction is responding to that evolution.”

Even vocal critics of the society's opposition to open-access publishing aren't delighting in its financial woes. Peter Murray Rust of the University of Cambridge, UK, whose blog covers open-access chemical information, says that he wishes the society well. “I have not been a supporter of many of [its] policies,” he says, “but I would say that we absolutely need national scientific societies.”

Other societies are also seeing the value of their portfolios decline. “Our investments fared about as well as anyone's,” says Mark Hernick, director of finance at the American Geophysical Union, also based in Washington DC. “Most investments lost 20 to 40% of their value last year alone, and we were in that category.” He says that his society is not planning staffing reductions.

Richard Pike, chief executive of the London-based Royal Society of Chemistry, says, “Our reserves are hit, but we are in a somewhat better condition than some others — we had a good year ending in 2008”. The UK society is actually adding staff, Pike says, mostly in its publications and membership divisions.

Ruskin says that he hopes ACS members won't notice any change in services provided by the society. “All the things that we provide to them are considered essential,” he says. “How do we continue to provide that in a manner that is sustainable?”

Emma Marris



Hit by the downturn and the online revolution.

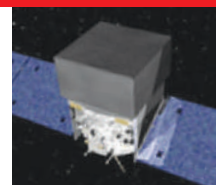


**FLU MONITORING**  
Zoos help track the spread of pandemics.  
[www.nature.com/news](http://www.nature.com/news)

PHOTODISC/PUNCHSTOCK

K. SIMPSON





**DARK MATTER**  
Experiments may have glimpsed mystery particles.  
[www.nature.com/news](http://www.nature.com/news)

NASA

# Neuroscientists claim growing pains

Manufacturer denies that common lab feed can cause some neuron cultures to fail.

Leading neuroscientists are warning that difficulties with a staple laboratory product may be costing time and money.

The scientists say that variation between batches of a growth medium designed to sustain neurons in culture can, in their experience, cause experiments to fail or give low-quality results because of the poor survival and maturation of cells.

The growth medium in question is a particular formulation of B27, a mixture of proteins, hormones and vitamins, produced by laboratory supplies company Invitrogen of Carlsbad, California, now a division of Life Technologies. In 2004, a group of neuroscientists including Beth Stevens, now at Harvard Medical School, in Boston, Massachusetts, and Johannes Hell of the University of Iowa, Iowa City, told Invitrogen that they thought the B27 medium was producing poor results following a change in the product's ingredients. The company says that it remedied the problem, and that the number of complaints it received from scientists fell to "negligible levels".

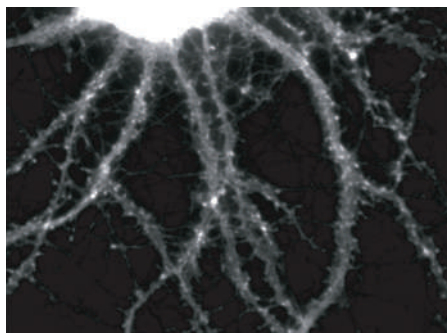
"Things were improved," says Stevens, "but over the next few years it became clear there is still a lot of variation in quality control." Hell estimates that his lab spends about US\$50,000 per year on culturing neurons. "If you get mediocre cultures, that money is wasted," he says.

So Hell and other colleagues developed an alternative growth medium called Neuronal Supplement 21 (NS21), which they unveiled last year (Y. Chen *et al. J. Neurosci. Meth.* 171, 239–247; 2008). Hell points out that NS21 is not a commercial product; the researchers who use it prepare it in their own laboratories.

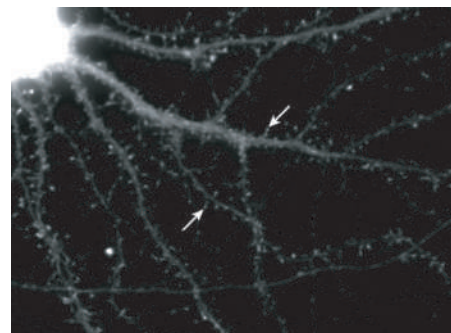
Charles Harata, a colleague of Hell at the University of Iowa who was not involved in the development of NS21, used B27 while working as a postdoc for Richard Tsien at Stanford University, California. Harata felt that the quality of his hippocampal-neuron cultures declined markedly in the years before he left Tsien's lab in 2007.

That summer he took up a full-time post at Iowa, initially using B27. About six months ago, however, he decided to try his new colleague Hell's NS21 formulation.

Harata conducted a blind comparison of B27 and NS21 to grow two neuron cultures in otherwise identical conditions. He found that B27 did not support the neurons as well as



Culture differences: dendritic spines (arrowed) are signs of development that were visible after neurons were cultured for 18 days in NS21 (right), but not in B27 (left), in Johannes Hell's laboratory.



Y. CHEN &amp; J. W. HELL

NS21, particularly at lower cell densities. "That was a clear effect for me," says Harata. "I've now completely switched to using NS21."

Hell says that in his capacity as a reviewer he has seen academic papers in which he suspects poor-quality cultures may be due to problems with B27. Oliver Schlüter, of the European Neuroscience Institute in Göttingen, Germany, agrees that there is "batch-to-batch variation" in B27. In his own lab, "we test the batches and compare them with what we had previously", Schlüter says. "If it behaves similarly we buy as much as we can."

**"If you get mediocre cultures, money is wasted."**

Joydeep Goswami, vice-president of primary and stem-cell systems at Life Technologies, says that batches of B27 can vary because the medium includes ingredients from biological sources, such as bovine serum albumin, that can be difficult to standardize. He adds that each batch goes through a rigorous quality-control process, assessed periodically by the US Food and Drug Administration, and "the variability should not affect customers that are following standard protocols".

Goswami asserts forcefully that B27 is not wasting researchers' time and money, adding, "We stand behind our products."

He suggests that variability in results may arise not from B27 itself but from the different types of cell grown in the medium, or the different procedures used by scientists running the experiments.

The neuroscientist who originally developed B27, Gregory Brewer, works at Southern Illinois University, Springfield, and is a consultant for Invitrogen. He points out that his original studies tested B27 on its ability to sustain neurons for four days. Now, he says, neuroscientists

often expect the medium to keep neurons in shape for weeks, the time period after which features such as synapses and dendritic spines appear (see pictures of hippocampal neurons from rat embryos, above).

Brewer still uses B27 himself, and intends to discuss with Invitrogen an assessment of whether the medium can support extended growth times.

Richard Haigh, director of product development for Invitrogen, agrees that B27 "is now used in a broad range of applications beyond its original purpose" and says that the company is looking at products that might better support researchers' specific cell types.

## Replication challenge

The precise makeup of Invitrogen's B27 is commercially sensitive, although the ingredients of the original B27 formulation are listed in the research paper that first defined the medium (G. Brewer *et al. J. Neurosci. Res.* 35, 567–576; 1993). "My biggest concern is that we're dealing with a black-box chemical," says Schlüter. "If you want to replicate other people's experiments, you need to know what's in there and what variation could happen."

Goswami points out that customers can get technical help from Invitrogen and can request data sheets for each batch.

Yet Stevens believes that scientists having problems with their neuron cultures may not even question if their growth medium could be a factor. "I'm quite sure people are banging their heads against a wall. Trying to troubleshoot this can waste months," she says.

"Frankly we don't hear the same thing," counters Goswami. "More than 1,000 labs are using B27 successfully."

**Daniel Cressey**

# Geologists suffer observatory glitches

Flagship experiment on the San Andreas fault has been troubled since last autumn.

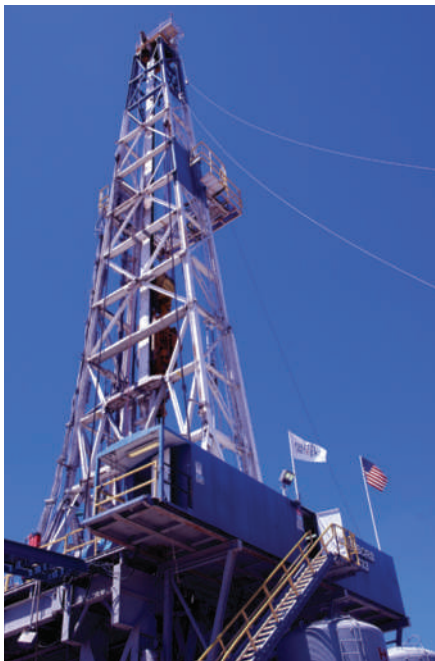
As US geophysicists gathered last week to celebrate EarthScope, one of their most ambitious programmes ever, researchers let slip an embarrassing fact that they had kept largely under wraps for 6 months. One major element of the project — a suite of instruments buried deep in California's San Andreas fault — is broken. Researchers are making do with a trickle of data from a temporary instrument.

"Unfortunately, the observatory stopped working a few days after it was installed" last September, says Mark Zoback, a geophysicist at Stanford University in Palo Alto, California, who was formerly one of three principal investigators on the project. "That was a big disappointment." There are, as yet, no firm plans or funds committed to fix the instruments.

The difficulties may serve as a warning to other countries, such as Japan, Greece, Italy and New Zealand, which are considering drilling into fault zones.

The problems occurred at the San Andreas Fault Observatory at Depth (SAFOD), located near the town of Parkfield about halfway between Los Angeles and San Francisco. The US\$25-million project is part of the \$200-million EarthScope effort, which in September ended its 5-year-long construction phase.

Researchers came to celebrate that milestone last week in Washington DC, and the praise was overflowing. "We are extremely proud of what



Deep trouble: a drill rig or crane could fix SAFOD.

you have achieved," said Robert Detrick, director of the division of earth sciences at the National Science Foundation, which funds EarthScope. "We're going to rewrite the textbooks on North American structure and dynamics."

The talks touched on a feast of data coming from the three components of EarthScope: a giant network of seismometers that will survey the entire United States; a set of 1,100 Global Positioning System stations, plus other instruments, tracking movement along North America's western edge; and SAFOD. That third element has delivered some impressive results, including the first fault rock taken at depth from a seismically active zone, but it has also caused massive headaches.

SAFOD sits on the San Andreas fault, where the Pacific and North American tectonic plates creep past each other. In 2004, Zoback and his colleagues drilled a vertical shaft near the fault, and the next year they curved the shaft to pass through the fault zone. In 2007, the team struggled to drill cores of rock from the fault itself. After weeks of drilling, with time and money running out, the team eventually succeeded in hauling up pieces of the fault at 4:00 a.m. local time, as a huge lightning storm provided a dramatic backdrop.

Last September, researchers lowered a set of nine instruments more than 3,100 metres down, right above a moving section of the fault. The package included three seismometers, three accelerometers (which measure strong motion), two tiltmeters for detecting shifts in the rock's orientation and an electromagnetic coil for picking up any electrical or magnetic

EARTHSCOPE

## UK scientists get funding ban reprieve

After a campaign by scientists, the UK Engineering and Physical Sciences Research Council (EPSRC) has softened and delayed its controversial policy to bar serially unsuccessful grant applicants from making funding bids for one year.

The ban — which may be unique among European and US funding bodies — was due to be imposed on 229 researchers starting on 1 June, in an effort to reduce pressure on an overloaded system that currently peer-reviews all grant applications.

But eight weeks after it published the policy (see *Nature* 458, 391; 2009), the EPSRC now says that the restriction will not come in until 1 April 2010 — giving scientists

more time to change their grant-submission behaviour so that they do not fall under criteria defining repeated failure. And instead of being excluded outright, researchers will be allowed one application during the year.

"We have made these adjustments to address concerns raised by the community — for example, the retrospective nature of [the policy's] implementation," the EPSRC said in a statement. "We've made bold changes to protect peer review, but we're not an insensitive organization."

Peter Main, director of education and science at the Institute of Physics in London, says the EPSRC

has listened to criticism and has shown flexibility. "It's the policy that it perhaps should have been in the first place," says Joe Sweeney, an organic chemist at the University of Reading, UK, who set up an online petition demanding the policy be repealed, signed by more than 1,900 scientists.

But some researchers say they are disappointed not to have been consulted more directly beforehand — which might have prevented the EPSRC from introducing the ban in the first place. "It's something of a shame that we had to force them into this policy change," says Philip Moriarty, a physicist at the

University of Nottingham, UK.

The EPSRC is keeping a policy introduced on 1 April, to refuse uninvited resubmissions of failed proposals, which it says will cut 20% of applications submitted for review. The exclusion policy had been expected to cut a further 10%.

The EPSRC says that letters intended to warn individuals in April were never sent. "We are an organization that listens to the community," says chief executive David Delpy. "If we can make amendments to help researchers whilst ensuring the overall policy is still effective, then that's in everyone's interest."

Richard Van Noorden



**HAVE YOUR SAY**

Comment on any of our  
News stories, online.

[www.nature.com/news](http://www.nature.com/news)

changes that might precede an earthquake.

The plan was to leave those instruments in the hole to capture the many small repeating earthquakes that occur within a few hundred metres. For a few days, the devices worked well. But then they went quiet. William Ellsworth of the US Geological Survey in Menlo Park, California, another former principal investigator, says there may be an

electrical short in one of the connections that carries data to the surface. Hot, corrosive groundwater in the hole could have damaged equipment.

Because the instruments are mounted on thick pipe, it would take a crane or a drill rig to pull the package back up. Right now, the project does not have the money to do that. In December, the team lowered one temporary

seismometer into the drill hole and is currently collecting data from that.

Ellsworth says some of the problems were discussed in December at a meeting of the American Geophysical Union in San Francisco. But there has been no general announcement to the public. "We were all embarrassed about it," says Zoback.

**Richard Monastersky**

# Hubble: the last hurrah

On 11 May, the space shuttle Atlantis will head to the Hubble Space Telescope for a fifth and final servicing mission. With six batteries, six gyroscopes and a data interface system to be installed by astronauts, the telescope should be ready to perform for at least another five years. For scientists, the main feast will come with the new Wide Field Camera 3 and the Cosmic Origins Spectrograph, along with repairs to existing instruments. "I can hardly wait," says David Leckrone, senior project scientist for Hubble at NASA's Goddard Space Flight Center in Greenbelt, Maryland. Eric Hand looks at some of the most highly ranked proposals from outside scientists.

## SPACE TELESCOPE IMAGING SPECTROGRAPH (STIS)

The repaired STIS can take observation after observation across objects such as galaxies and black holes. It is also good for detecting chemicals in the atmospheres of transiting exoplanets. David Sing of the Astrophysics Institute of Paris wants to look for sodium in the atmosphere of a known 'hot Jupiter'. The measurements would provide stringent constraints on models that describe the exoplanet's atmosphere.

## COSMIC ORIGINS SPECTROGRAPH (COS)

COS will be very sensitive to far ultraviolet light and well suited to splitting up the light created by point sources, such as quasars and stars. Absorption features in this light reveal details of the clouds of gas the light has passed through. At the moment, astronomers can't find all the gas that they expect. Todd Tripp of the University of Massachusetts in Amherst will use COS to look for cool gas clouds backlit by quasars.

## FINE GUIDANCE SENSOR (FGS)

Hubble uses three fine guidance sensors to keep track of where it points. These can also be used to look for wobbles associated with planets. One of its three sensors will be replaced.

## NEAR INFRARED CAMERA AND MULTI OBJECT SPECTROMETER (NICMOS)


NICMOS, an actively cooled instrument in the near infrared, has been offline since September, and several reboot attempts haven't worked. Engineers suspect that an ice particle in its cryocooler could be causing problems.

## WIDE FIELD CAMERA 3 (WFC3)

WFC3 will be by far the most popular instrument, and Leckrone says an experiment proposed by Garth Illingworth, an astronomer at the University of California, Santa Cruz, is the most highly ranked of all. Illingworth wants to use WFC3 to stare in the same spot as the Hubble Ultra Deep Field, a million-second exposure that captured 10,000 galaxies in one of the longest stares back at the beginning of time. In his 'ultra ultra deep field' image, Illingworth expects to find 100 objects as old as 13 billion years — providing the statistical strength needed to understand the timing and mechanism behind the build up of galaxies.

## ADVANCED CAMERA FOR SURVEYS (ACS)

ACS, installed in 2002, is similar to WFC3 but slightly more sensitive to faintly red objects, such as the dwarf galaxies that Kristin Chiboucas, of the University of Hawaii in Honolulu, wants to find. Based on models for galaxy formation, an order of magnitude more dwarf galaxies are expected to be clustered near big galaxies such as the Milky Way than have been found. Chiboucas hopes to complete a census of the dwarfs surrounding the nearby spiral galaxy M81.

 Indicates the percentage of observing time spent on each instrument.



## Australia delays carbon reduction scheme

The Australian government has delayed its proposed cap-and-trade scheme for regulating greenhouse-gas emissions (see *Nature* 458, 554–555; 2009).

Prime Minister Kevin Rudd announced on 4 May that, in order to “manage the impacts of the global recession”, the system will be phased in from 1 July 2011, a year later than planned. Until July 2012, permits to emit carbon dioxide will be sold at a fixed price of Aus\$10 (US\$7.5) per tonne, and companies involved will be able to buy an unlimited number of them.

Rudd said he hopes to push the necessary legislation through parliament this year. He will also revise the upper limit on Australia's emissions-reduction target to 25% below 2000 levels by 2020, up from 15%, depending on agreements reached at December's UN summit in Copenhagen.

## Data ‘mishandling’ stalls Down's syndrome test

Research data supporting a non-invasive prenatal screen for Down's syndrome were “mishandled” and cannot be relied on, according to the biotechnology company that has been developing the test.

On 29 April Harry Stylli, chief executive of Sequenom in San Diego, California, said that four employees suspected of mishandling data had been suspended. An independent committee is investigating the matter. Investors have since filed several class-action lawsuits against the company.

Sequenom's tests rely on examining minute amounts of fetal DNA or RNA circulating in maternal blood. The hope is to provide an alternative to more invasive methods such as amniocentesis, which carry a small risk of miscarriage. The company intends to complete a new clinical validation of its assays by the end of the year, but will not file for regulatory approval until two larger clinical trials are completed in 2010.

## Misconduct scandal hits German university

Sixteen members of a Collaborative Research Centre (SFB) at the elite University of Göttingen, funded by Germany's major national research agency, the DFG, are being investigated for scientific misconduct.

SFBs are highly prized and generously funded awards that continue for up to 12 years. In a statement on 4 May, the university said that an internal preliminary investigation had found that the scientists listed non-existent papers in their routine

## Obama revokes endangered species rule

President Barack Obama last week revoked a rule that limited the role of wildlife experts in reviewing roads, pipelines and other projects under the Endangered Species Act.

The rule, established in the last weeks of George W. Bush's administration, allowed federal agencies planning such projects to make decisions about whether threatened or endangered species would be harmed, without first consulting the National Marine Fisheries Service or the Fish and Wildlife Service. Among species currently being considered by the Fish and Wildlife Service for protection under the Endangered Species Act is the American pika (pictured).

The Bush administration said that the change reduced unnecessary paperwork, but environmentalists said that it undermined a key safeguard.

Obama had promised to review the issue in early March, and the Department of the Interior followed up by rescinding the rule on 28 April.



S. HARADA/MINDEN PICTURES/FLPA

milestone report for continuation of funding. The researchers have apologized to the DFG for providing false information.

The interdisciplinary project, studying the stability of the Indonesian rainforest, had been approved €16.6 million (US\$22.3 million) from 2000 to June this year. A request for a further €8.6 million has been withdrawn. The DFG is looking into the case and investigating whether financial irregularities occurred.

## South Korea re-enters human stem-cell research

On 29 April, South Korea's national bioethics committee conditionally approved a study using human eggs for stem-cell studies — the first such licence granted since a scandal unfolded over similar work by Korean researcher Woo Suk Hwang.

A team at Cha General Hospital in Seoul will be permitted to create cloned human embryos using the eggs. The study's go ahead is dependent on four conditions: donors must give informed consent; research with human eggs should be

minimized; an internal review board must oversee the experiments; and the study cannot be hyped by mention of possible clinical benefit.

The research aims to establish stem-cell lines from the cloned embryos, for research and potentially for therapy. Hwang published claims to have done that in 2005, but in January 2006 his research was found to be fraudulent and his acquisition of human eggs unethical. He remains on trial for fraud, embezzlement and violation of the nation's bioethics law.

## Biosecurity report cautious on strict researcher vetting

Academic researchers working with dangerous pathogens should not be subjected to the same stringent background checks currently required for scientists working in chemical and nuclear weapons programmes, a panel of US government advisers recommended last week.

The National Science Advisory Board for Biosecurity report, released on 29 April, examined how to reduce “insider threats” posed by bioscientists with malicious intent. It noted that security regulations governing bioscientists were bolstered after the US anthrax attacks of 2001, and said they should continue to be strengthened.

But piling on “onerous” requirements, it said, could cause the best scientists to shun important work with pathogens and could isolate those that remain in the field.

A working group, established by former president George W. Bush in January, is reviewing US lab biosecurity and is expected to report in July.



Woo Suk Hwang: still on trial for stem-cell fraud.

# STRUCTURES OF DESIRE

What do protein crystallographers dream of? The eukaryotic ribosome, the spliceosome, the nuclear-pore complex, the HIV trimer and almost any transmembrane protein, finds **Ananyo Bhattacharya**.

**W**hen considered up close, the blood protein from a sperm whale is a marvellous thing. Or so it seemed just over 50 years ago, when John Kendrew and other researchers at the Cavendish Laboratory in Cambridge, UK, reported that they had used X-rays to reveal the three-dimensional structure of a globular protein for the first time. Analysis of the diffraction pattern caused by crystals of myoglobin, which was chosen for its simplicity, required one of the most powerful computers in the world at that time, and later won Kendrew a share of the Nobel Prize in Chemistry with his Cavendish colleague Max Perutz. The picture it created “is more complicated than has been predicated by any theory of protein structure”, Kendrew and his colleagues wrote in a *Nature* article<sup>1</sup>.

Half a century on, X-ray crystallography's techniques are in outline the same: you need a crystal, X-rays and calculating power to make sense of the diffraction pattern. In all these three areas, however, progress has been

enormous. Washing machines now have more computing power than the computers Kendrew used and synchrotrons offer X-rays a trillion times more brilliant than those available in Cambridge 50 years ago. Growing crystals is increasingly routine: with many proteins it can be an automatic process, and the skills needed for the harder ones have come on in leaps and bounds.

But so, too, have the ambitions of the crystallographers. They still believe, as did those earliest molecular biologists, that there is no better way to understand a complicated machine than to capture an atomic-scale picture of it. But now they dream of using their techniques on things far more challenging and complex than anyone imagined 50 years ago — including some machineries of protein and nucleic acid that dwarf myoglobin as a whale does a minnow.

## THE BUILDING SITE

The crystals look beautiful under the microscope. Suspended in a tiny drop of solvent, each gem-like growth is a symmetrical array of copies of one of the two subunits of **the ribosome** — the cell's machine for making proteins. Each of those subunits is a tangle of many proteins and RNA. What makes the crystals especially interesting is that their constituents come not from bacteria but from eukaryotic cells. Which plant, animal or other eukaryote provided those cells, though, is for the moment a secret.

Such crystals are rare and highly sought after. Unfortunately, Nenad Ban ruefully admits, these ones grown in his lab at the Swiss Federal Institute of Technology in Zurich (see photo, right) will not be yielding a protein structure any time soon. A crystal that looks pretty in visible light can still be a mess by the demanding standards of X-rays. “We have fantastic-looking crystals,” he says. “But they are not very good in terms of diffraction.”

Ban has been here before. Nearly a decade ago he was in one of several teams striving to

solve the structure of the simpler ribosomes used by bacteria and archaea. He and his colleagues in Tom Steitz's lab at Yale University published the structure of the 50S subunit of the ribosome of *Haloarcula marismortui* — a microorganism that lives in the famously salty waters of the Dead Sea — in 2000 (see ref. 2). The 50S subunit consists of about 30 proteins and has a mass of 1.5 million daltons, compared with myoglobin's 1,700 daltons or a carbon atom's lowly 12. Just a month later, a second team published the structure of the smaller 30S ribosomal subunit, using material from the bacterium *Thermus thermophilus*<sup>3</sup>. The following year, Harry Noller at the University of California, Santa Cruz, and his colleagues unveiled the structure of the whole bacterial ribosome, revealing much about how it binds to the transfer RNAs that deliver amino acids to a growing protein<sup>4</sup>.

These structures triggered an avalanche of new work in a field that had, in their absence, largely ground to a halt. They allowed researchers to see, for example, how the ribosome

catalyses the joining of one amino acid to another with a peptide bond. Now, says Jennifer Doudna of the University of California, Berkeley, “we know a lot of detail about what the bacterial ribosome looks like, how it works, how peptide bonds are made, and even a lot about how the initiation process is regulated in bacteria”.

The eukaryotic ribosome takes the competition to another level. It is bigger — containing some 80 component proteins compared with the bacterium's 50 to 60 — but that is not the only, or even the main, challenge. “There's a lot more bells and whistles; there's a lot more regulation that goes on,” Doudna says.

It's those bells and whistles that make the eukaryotic ribosome more difficult to work with than its bacterial counterpart. In mammals, for example, a host of additional proteins called initiation factors interact with the ribosome. The initiation factors are themselves



## THE EDITING SUITE

Noller, who with colleagues cracked the structure of the complete bacterial ribosome, says the next big thing is another signature speciality of the eukaryote, **the spliceosome**. “The spliceosome would be fantastic,” he says. “That would make the ribosome look like child’s play.”

Made up of 150 or so proteins, the spliceosome slices and dices freshly made messenger RNA, stitching together the ‘exon’ sequences that will be translated into protein and relegating the ‘introns’ to the cell’s cutting-room floor. It can bring together sections of RNA maybe tens of thousands of base pairs apart and then snip out the intervening loop, like a movie editor running through many metres of film to splice two shots together. The activity of the spliceosome fascinates biologists because it could help explain how eukaryotic cells generate biological complexity — in the form of different RNAs and proteins — from a single DNA sequence. But that fast and continuous activity is also what makes it a gruelling challenge for crystallographers.

“Tremendous work has been done genetically and biochemically to understand how splicing works and how it’s regulated,” says Doudna. “The big missing piece in that field is not having access to high-resolution structural information for how the spliceosome is put together and what is driving the ... changes that have to occur during the splicing process.”

The problem for crystallographers is that the spliceosome is not just one machine — it is five, and all are in constant motion. Called small nuclear ribonucleoprotein particles (snRNPs, pronounced ‘snurps’), these five protein–RNA complexes come together transiently in a complicated, fast-moving dance, their

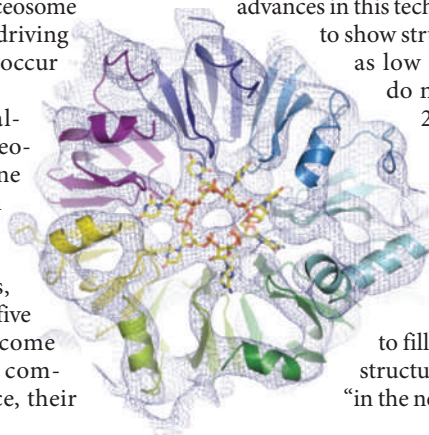
mercurial assembly about the size of a ribosome but far less stable. “Unlike the ribosome, you cannot simply purify a spliceosome from cellular extracts because you have many different spliceosomal complexes — different snapshots at different stages of function,” explains Reinhard Lührmann of the Max Planck Institute for Biophysical Chemistry in Göttingen, Germany, whose group has crystallized several proteins that are components of snRNPs.

Stalling the spliceosome at a particular stage of the cycle is a “major challenge” says Lührmann. One way might be to use a mutant RNA message or a small molecule to arrest the spliceosome mid-splice.

A team led by Kiyoshi Nagai at the Medical Research Council Laboratory of Molecular Biology in Cambridge, UK, recently published a relatively low-resolution version of the crystal structure of one of the smaller snRNPs, which they reconstituted from its RNA and seven recombinant proteins (see graphic below)<sup>5</sup>. Reconstituting all 150 proteins in the complete spliceosome, however, “does not appear feasible” in the foreseeable future, Lührmann says.

Instead, he is pursuing the structure using cryo electron microscopy, in which samples are flash-frozen in liquid ethane to protect them from the bombardment of high-energy electrons in an electron microscope. Recent advances in this technique have allowed it

to show structures at resolutions as low as 5 angstroms, but do not yet approach the 2-angstrom resolution of good X-ray crystallography. Lührmann predicts that further advances in hardware and software will allow cryo electron microscopy to fill out the broad atomic structure of the spliceosome “in the next few years”.



M. MUELLER/THE BAN LAB

REF. 5

complex assemblies — eukaryotic initiation factor 3 (eIF3), for example, is made up of at least 12 proteins and is only a few times smaller than the ribosome itself. Ribosomes that are purified from a cell could be in any number of combinations with these and other proteins.

Getting material that is pure and homogeneous is a significant hurdle, says Doudna. Simpler proteins can be mass-produced by inserting the appropriate gene into a cell culture that then churns out proteins, but ribosomes are too large and complex to be produced in this way.

Some who worked on the bacterial structures are joining the hunt for the eukaryotic one, and the competition is fierce. That is why Ban will not reveal from which eukaryote his ribosomes are harvested, nor which subunit his lab has managed to crystallize. The fact that he has crystals — albeit ones that don’t diffract X-rays well — is an important proof of principle, he says. But “it’s the endgame that counts”.

## THE MONSTROUS MAW

More than 30 times the mass of the ribosome and around 100 nanometres wide, **the nuclear-pore complex** is a doughnut-shaped assembly of hundreds of proteins that straddles the eukaryotic nuclear membrane. One of the largest protein conglomerations in the cell, the structure serves as both gate and gatekeeper, choosing which nucleic acids, proteins and other molecules to let in and out of the nucleus.

With only around 200 pores in a yeast cell —

compared with perhaps 10,000 to 20,000 ribosomes in a bacterium — purifying the complex from cells is “really impossible”, says André Hoelz, a researcher who works on the nuclear-pore complex in Günter Blobel’s laboratory at The Rockefeller University in New York.

Hoelz and his colleagues are taking a different tack — expressing and crystallizing single proteins or small protein complexes from the pore, and then piecing them together like a

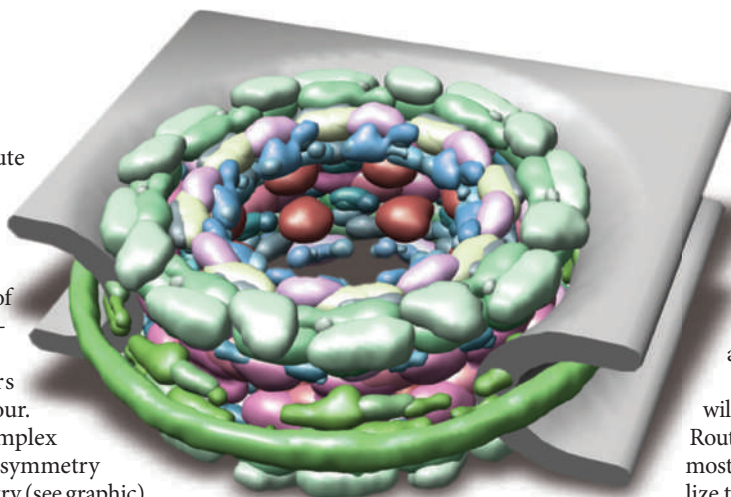


jigsaw puzzle to reconstitute the whole structure. “When we started this work five years ago people were saying that you can’t possibly get this done because of the sheer size of the nuclear-pore complex,” Hoelz says.

There are two factors working in the group’s favour. First, the nuclear-pore complex has an eightfold rotational symmetry and twofold mirror symmetry (see graphic). That means that the hundreds of proteins that make up the pore are in fact made up of repetitive arrangements of only about 30 different types — around half the variety in the ribosome. Second, in many organisms the nuclear-pore complex is dismantled when cells break down their nuclear membrane before dividing, and later reassembled piece by piece. It is always broken into the same building blocks, and it is these conserved components that Hoelz hopes to crystallize and slot together. If this approach is successful, it will provide a detailed ‘pseudo-atomic’ picture of the nuclear pore’s structure, although it may not provide the clarity that a crystal structure of the entire complex would.

There’s another feature of the nuclear-pore complex that is a challenge for crystallographers. The centre of the pore is a mesh of fluttering protein filaments that don’t fold in a regular way, but instead flop and dangle; they are thought to play a crucial role in selecting which proteins are ferried through the pore. Because a crystal structure is really an average of the arrangements of the atoms in millions of protein molecules in the crystal, the tentacles, which are in constant motion, would be an ill-defined blur. “Imagine one could crystallize the nuclear-pore complex; a quarter of it would be natively unfolded, and that’s the business end,” says Michael Rout, whose Rockefeller University lab also studies nuclear-pore complexes. So far, only isolated structures for some bits of the tentacles exist, solved by teams that break them off and crystallize them separately.

The nuclear-pore complex, then, runs up against a fundamental limit of crystallography — it generates snapshots, not movies. And it is not alone: up to 30% of eukaryote proteins are wholly or partly disordered. To see proteins in action, some crystallographers and modellers have turned to computer simulations that jump between two or more ‘frames’, each obtained by crystallography. Wayne Hendrickson of Columbia University in New York says there is also a lot of excitement about technologies that might be possible at facilities such as the European XFEL (X-Ray Free Electron Laser) under construction



in Hamburg, Germany (see page 16). The idea here is that an extremely short burst of X-rays could be scattered off a single protein molecule, blowing it apart but revealing something about its structure before the disintegration. “You are in principle able to capture the molecule in action,” Hendrickson says.

That’s something for the future: the XFEL will not be commissioned until 2014. For now, Rout says, “the current approach is probably the most successful, which is to continue to crystallize the structured parts and put that together with other data to build a complete picture”.

## THE KILLER KEY

“A big missing piece in the virology field is the structure of **the HIV trimer**,” says Ian Wilson, a structural biologist at the Scripps Research Institute in La Jolla, California. “We don’t understand what that looks like.” It’s a challenge that Wilson and his colleague Robert Pejchal have taken on relatively recently — though they’re not the first. “There have been many people in and out of the game for years because it has been so challenging.”

The trimers that Wilson and others want are protrusions from the surface of HIV, also called ‘envelope spikes’. Each one has a tripartite structure: three gp41 proteins rise from the viral envelope, forming a stem that supports three gp120 molecules. It has long been thought that when the spike binds to key receptors on white blood cells it triggers massive structural changes in the trimer that drive the fusion of virus and cell.

Trimers are the focus of intense vaccine research efforts, but vaccines based on the trimer so far do not stimulate enough of an antibody response to combat a later infection (see *Nature* 454, 565–569; 2008). “There’s something about the trimer that makes it difficult to mount an effective immune response,” Wilson says. If they had the trimer’s structure to work from, researchers hope they could devise better ways to turn the human immune system against it.

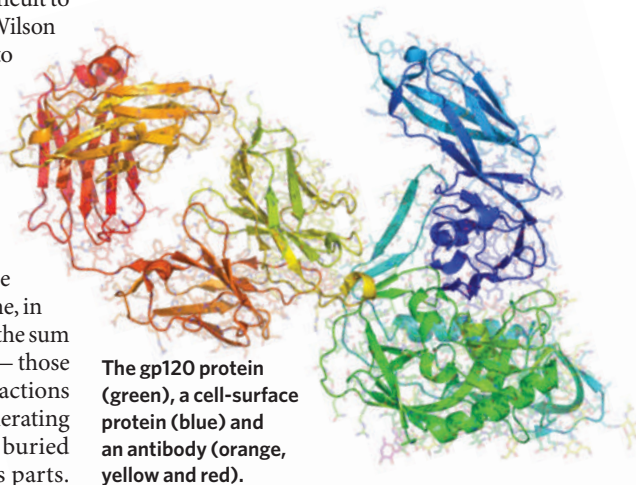
The problem has been that, isolated from the virus, the trimer falls apart. Researchers can’t adopt the nuclear-pore tactic here and break it apart before piecing it back together. The situation is more like that of the ribosome, in which the whole structure is more than the sum of its parts. Some regions of the trimer — those that are important for the virus’s interactions with cell-surface receptors and for generating an effective immune response — are buried inside or in the interface between its parts.

Researchers can’t tell what they normally look like unless they see a trimer intact.

The first structure of a pruned version of gp120, combined with a fragment of the white cells’ CD4 receptor and an antibody, was published in 1998<sup>6</sup>. Other structures of gp120 and gp41 have followed<sup>7–9</sup>. And a recent cryo-electron-microscopy study of whole virus particles revealed some of the large-scale changes that occur when the trimer binds to the CD4 receptor<sup>10</sup>. But the structure of the trimer proper has remained out of reach.

Wilson is making progress in collaboration with John Moore at Weill Cornell Medical College, New York, who has engineered a disulphide bond between gp41 and gp120 that helps to lock them together. They also used a strain of the virus that tends to form more stable trimers — called SOSIP trimers — in the first place. “We have SOSIP trimers and we have crystals, but they don’t diffract well,” says Wilson.

Still, Wilson and others will continue to plug away at the problem — because a vaccine remains “one of the big challenges to science in the present day”, says Hendrickson.



The gp120 protein (green), a cell-surface protein (blue) and an antibody (orange, yellow and red).

## BEST OF THE REST: MORE STRUCTURES HOTLY DESIRED

### Internal ribosome entry site

A sequence of RNA that allows ribosomes to begin protein production from the middle. Helps RNA-based viruses to hijack protein production.

The sequence is very flexible. Parts have been crystallized — but not the whole thing.

### Sec61 (eukaryotic translocon)

A protein complex that can help

an unfolded polypeptide chain to cross or span a membrane. Recalcitrant when it comes to crystallization.

### Telomerase RNP

Adds repeating sequences of DNA to the end of chromosomes and plays a role in cancer and ageing. Individual components of the telomerase have been solved.

It's a flexible, hard-to-get

material and it interacts with other molecules.

### Higher-order chromatin structure

The complex of protein and RNA with which DNA is packaged. The core of a single nucleosome — the basic subunit of chromatin — has been solved to high resolution.

Crystallographers want higher-order structures of chromatin — two or preferably more

nucleosomes. This is proving difficult.

### Sodium-selective ion channel

Ion channels that allow sodium ions to pass but not potassium ions — essential in generating electrical potentials in neurons.

They are formed by one big polypeptide chain, which may make them harder to crystallize than potassium channels.

## THE INVISIBLE THREAD

Stephen White's website at the University of California, Irvine, features an exponential curve of which structural biologists are proud. In 1985, it shows, the first structure of a membrane protein was solved<sup>11</sup>. (That work, on the photosynthetic reaction centre, won a Nobel prize.) Now, the number of structures for proteins that span membranes deposited in the Protein Data Bank has climbed to more than 180.

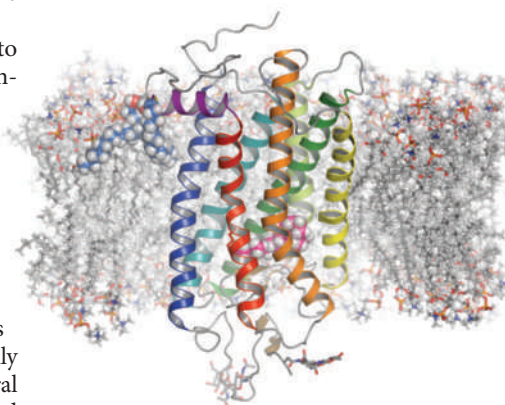
Some membrane proteins are refusing to join the cavalcade, however. Take, for example, **the epidermal growth factor receptor** — the target of Genentech's breast cancer drug Herceptin (trastuzumab). Despite a decade or more of intense study, only the bits protruding outside and inside the cell membrane have been crystallized. The connecting portion — the bit that spans the membrane and transmits information from one side to the other — has not. The same is true for the 60 or so other proteins in this family of receptor tyrosine kinases, which have central roles in cell proliferation, differentiation and disease. Solving this delicate stretch of protein would begin to explain how a signal outside the cell — such as a growth factor, hormone or other 'ligand' that binds to the receptor — can cause a change in protein conformation that leads to a response inside the cell.

To crystallize a membrane protein, you have to ease it out of its normal milieu. Released from their membranes, though, the proteins easily lose their shape and precipitate out of solution. So detergents are used to keep them soluble, folded and active, and these can end up being a problem themselves. Unless just the right types and amounts of detergents can be found, their molecules can obstruct the interactions between proteins that allow them to line up and form crystals.

The particular problem for the receptor tyrosine kinases is not their bulk or complexity but

their flimsiness. The peptide chain that connects their extracellular and intracellular parts snakes through the membrane only once. The head or tail can wobble around on the single transmembrane stem and make it difficult for the protein to form ordered crystals.

Persistence has paid off when it comes to another major class of membrane proteins.



There was much jubilation among structural biologists when, in 2000, a group led by Masashi Miyano at the Riken Harima Institute in Japan crystallized rhodopsin, a light-activated protein purified from the cow retina (see graphic, showing the helices of the protein embedded in a membrane)<sup>12</sup>. It was the first structure to be resolved in the class of G-protein-coupled receptors, a family of membrane receptors with almost 1,000 members found in humans. Solving a second such receptor, which was done in 2007, was still a marathon task: one group set up 15,000 trials using a robot to optimize crystallization conditions<sup>13–15</sup>. These proteins have seven membrane-spanning regions, and may have been more tractable because this wider 'bridge' between head and tail stops their extremities wobbling around so much.

There's a long way to go yet, however, according to Hendrickson. "From my perspective," he says, "these structures haven't answered the questions that I want to answer, which are about how the activation process happens — how these proteins do their job when activated by a ligand."

A full-length receptor tyrosine kinase remains a dream structure, and one that many crystallographers doubt can ever be realized because there is no obvious way to stabilize the head and tail. "I'm not certain that's going to be feasible," says Hendrickson.

White says that membrane proteins in general, though, are getting less intimidating. "There are a lot more people with the courage to tackle membrane protein," he says. Their courage will grow as ways to make better crystals and brighter X-rays come online.

At the cutting edge, however, where crystallographers face the complexity of a nuclear pore or the wavering heart of a transmembrane protein, something extra is needed. "Know the protein intimately," White says. "So far, that seems to be a really important issue — to have somebody who loves the protein," he says. ■

**Ananyo Bhattacharya** is *Nature's* deputy news editor.

1. Kendrew, J. C. *et al. Nature* **181**, 662–666 (1958).
2. Ban, N., Nissen, P., Hansen, J., Moore, P. B. & Steitz, T. A. *Science* **289**, 905–920 (2000).
3. Wimberly, B. T. *et al. Nature* **407**, 327–339 (2000).
4. Yusupov, M. M. *et al. Science* **292**, 883–896 (2001).
5. Pomeranz Krummel, D. A., Oubridge, C., Leung, A. K. W., Li, J. & Nagai, K. *Nature* **458**, 475–481 (2009).
6. Kwong, P. D. *et al. Nature* **393**, 648–659 (1998).
7. Zhou, T. *et al. Nature* **445**, 732–737 (2007).
8. Huang, C.-C. *et al. Science* **310**, 1025–1028 (2005).
9. Chan, D. C., Fass, D., Berger, J. M. & Kim, P. S. *Cell* **89**, 263–273 (1997).
10. Liu, J., Bartsch, A., Borgnia, M. J., Sapiro, G. & Subramaniam, S. *Nature* **455**, 109–113 (2008).
11. Deisenhofer, J., Epp, O., Miki, K., Huber, R. & Michel, H. *Nature* **318**, 618–624 (1985).
12. Palczewski, K. *et al. Science* **289**, 739–745 (2000).
13. Cherezov, V. *et al. Science* **318**, 1258–1265 (2007).
14. Rosenbaum, D. M. *et al. Science* **318**, 1266–1273 (2007).
15. Rasmussen, S. G. F. *et al. Nature* **445**, 383–387 (2007).





# ENTER THE OXIDES

Thin films of oxygen-bearing compounds could have myriad practical applications, finds **Joerg Heber**, if a few problems can be overcome.

In late 1996, a young Bell Labs physicist named Harold Hwang told his lab director that he wanted to start a radical programme of research into oxides — the ubiquitous, oxygen-bearing compounds found in everything from granite and glass to ceramics, chalk and rust. Hwang was convinced that even the most familiar oxides might show surprising and useful properties if different ones could be stacked up into ‘heterostructures’: layer-cake-like arrangements in which each level is an ultrathin film just a few atoms thick.

The lab director, Horst Störmer, was slightly dubious — not about the potential, but about the practicality. “Have you ever grown a thin film in your life?” he asked. He knew all too well what Hwang was getting himself into. Störmer had made his own reputation by growing and studying thin films of a very different class of materials: semiconductors. Those films had shown some remarkable properties — including a phenomenon called the fractional quantum Hall effect, in which the free-roaming electrons inside a layer condense into a liquid-like state. That discovery would later earn Störmer a share of the 1998 Nobel Prize in Physics. But such phenomena appeared only if the layers were absolutely uniform in height, with a crystalline structure that was so pure and defect-free that electrons could race along without crashing

into imperfections. It had taken Störmer and his colleagues at Bell Labs more than 10 years to invent and perfect the techniques for fabricating such films. And oxides, he knew, would be even more difficult to master. The compounds, which form 99% of Earth’s outer crust, typically consist of a larger number of chemical elements than semiconductors, and have more complex crystal structures.

Still, Störmer told Hwang to go ahead, and the younger man did not disappoint. In time, Hwang and others doing research in the field succeeded in growing high-quality oxide thin films with the same atomic precision as semiconductors. And those films do indeed exhibit interesting phenomena. In 2004, for example, Hwang co-discovered the existence of a two-dimensional (2D) electron gas, in which electrons at the interface of two oxide thin films show an extremely high mobility<sup>1</sup> — an effect that is particularly striking because the two oxides involved are electrical insulators.

Now oxide thin films are at roughly the same stage of development as semiconductor thin films were in the early 1970s — a period when researchers were finally learning how to work with them well enough to fabricate devices

such as the thin-film lasers, which would later have their commercial breakthrough in compact-disc players. For example, the 2D electron gas that Hwang and his colleagues discovered is being explored for use in a new type of fast transistor, a device that can amplify or switch electronic signals. Another use of oxide films could be as the basis for very high-density data-storage devices in which the magnetic information is controlled with electrical fields.

And that’s just the beginning, says Hwang.

“The great opportunity we have now is to design and grow artificial thin film structures down to the atomic scale — using multilayers of superconductors, ferromagnets, or even a combination — and to engineer systems that may one day be used for electronics or sensing applications.”

**“Have you ever grown a thin film in your life?”**

— Horst Störmer

## The power of oxygen

The rich array of phenomena found in oxides is largely due to the oxygen, says Yoshinori Tokura, a physicist from the University of Tokyo who has worked in this field for more than 20 years. Oxygen tends to pull electrons away from other atoms in the compound, says Tokura, resulting in strong electrical fields at the interatomic scale. These fields can give

D.G. SCHLÖM ET AL. J. AM. CERAM. SOC. 91, 2429–2454 (2008).



rise to substantial correlations in behaviour between the electrons of one atom and those of its neighbours. And the correlations in turn can lead to effects such as ferromagnetism, in which a material's electrons spontaneously line up and produce a magnetic field.

Nevertheless, for many years researchers tended to shy away from using oxides in advanced applications, because they are far more difficult to fabricate than metals and semiconductors. This situation changed in 1986 with the discovery of high-temperature superconductivity in certain oxides. The work kicked off an intense, worldwide focus on oxides that led to other discoveries. In 1993, for example, researchers encountered 'colossal magnetoresistance', in which a slight shift in the external magnetic field causes certain oxides to undergo an orders-of-magnitude change in electrical resistance<sup>2</sup>.

Another example is the 2D electron gas that Hwang and his co-worker Akira Ohtomo stumbled on when they were studying the interface between two insulators<sup>3</sup>, lanthanum aluminate ( $\text{LaAlO}_3$ ) and strontium titanate ( $\text{SrTiO}_3$ ). "We started to fabricate very crude-looking transistors that should not have been conducting by themselves, but found they were already conducting," recalls Hwang, who is now at the University of Tokyo. "We started thinking, 'What is going on here?'"

They soon found that everything depended on the precise crystalline structure of the interface: only when the right atomic layers met would the internal electrical fields on each side push electrons towards the junction, so that they could form the electron gas. Otherwise, no charge layer develops<sup>4</sup>. The interface electrons also turned out to be surprisingly mobile. In fact, as discovered in 2007 by the groups of Jochen Mannhart from the University of Augsburg in Germany and Jean-Marc Triscone from the University of Geneva in Switzerland<sup>5</sup>, these structures can become superconducting, meaning that the electrons can travel without resistance — albeit only below the very low temperature of about 200 millikelvin.

Researchers have also been studying potential applications that would exploit the thin-film interface. One way to do that would be to place a ferromagnetic oxide next to an insulating oxide that isn't ferromagnetic. If an external electrical field is applied, it causes an electrical polarization to develop at the interface. But the field also shifts the number of electrons in the ferromagnetic material, which changes the magnetic field. As a result, electrical polariza-

tion and magnetism are both controlled by the same electrical field, and are therefore cross-linked — a coupling of properties that defines multiferroic materials<sup>6</sup>. Such materials are of interest both as magnetic field sensors and as memory devices, in which information is written by electrical voltages and read by magnetic read head — with the benefit that no electrical current flows through the device, significantly reducing heat generation. Indeed, says Tokura, "the route via thin films offers the most straightforward fabrication method to realize a multiferroic material".

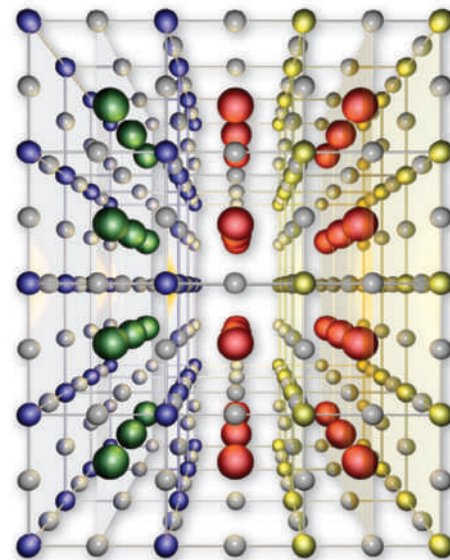
As well as looking at the coupling of two different properties at an oxide interface, researchers are looking for applications in which a single property, such as magnetic

field<sup>7,8</sup> or electrical conductivity, is controllably turned on and off. "Controlling conductivity as a whole, rather than electrical current itself, in some sense is the most exciting area," says Stuart Parkin, a physicist at IBM's Almaden Research Center in San Jose, California. "Contrary to conventional transistors, the required current densities could be quite small, and this is what you want for applications."

Consider, for example, the superconducting 2D electron gas. Through the application of an electrical field it is easy to push electrons away from the interface, destroying the superconducting state and making it impossible for current to flow<sup>9</sup>. This is analogous to what happens in conventional transistors, in which the flow of electrons can likewise be switched on or off by an external electrical field. Conceivably, researchers could use local voltages to

**"You can imagine controlling the states at phase boundaries by making small modifications."**

— Stuart Parkin



The interface between lanthanum aluminate and strontium titanate.

write complex patterns into this 2D electron gas. Where a voltage is applied, the interface would be insulating, and elsewhere it would be superconducting — potentially allowing the definition of entire electronic circuits. "It will be exciting to see the realization of small devices such as logic and memory circuits, or even small amplifiers," says Mannhart. Amplifiers written into the superconducting film could enable fast switching with extremely low noise levels and thus could detect and amplify weak electronic signals. Even the logic gates used for quantum information processing could be etched into the superconducting layer this way.

### Just one small push

Unfortunately, the switching of superconductivity in the  $\text{LaAlO}_3/\text{SrTiO}_3$  system occurs at temperatures far too low to be relevant for most applications. So one alternative is to look at the different phases many oxides show at various temperatures or pressures. Conductivity often changes dramatically at the transition from one phase to another. "If you go to phase boundaries, that's where you often get extremely large instabilities," says Parkin. "Then you can imagine controlling those states by small modifications." Such a small trigger impulse can push the system from one phase to the other. This is what happens in colossal magnetoresistance — a small external magnetic field induces huge variations in electrical resistance during a phase transition.

To realize high-quality oxide heterostructures for applications, researchers have had to overcome substantial obstacles to the development of suitable thin-film growth techniques. Oxides often have complex crystal structures,





and films refuse to grow properly unless the right crystal layer is exposed on the top surface. Otherwise, the incoming atoms will not be able to stick to the proper chemical bonds. In 1994, this problem was identified and solved for  $\text{SrTiO}_3$  by Masashi Kawasaki, a materials scientist then at the Tokyo Institute of Technology, now at Tohoku University in Japan, who developed a pre-growth treatment involving various acids that strip the crystal substrate down to the desired atomic layer<sup>10</sup>. With this advance, says Kawasaki, “people could finally grow complex oxides”.

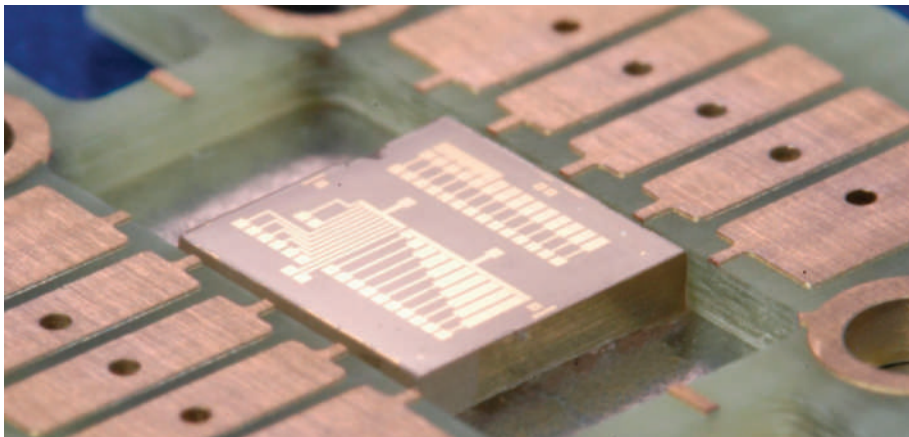
Unfortunately, it is still impossible to obtain oxide heterostructures anywhere near as large as those used in silicon technology. “Scaling up  $\text{SrTiO}_3$  wafers to realistic sizes is out of the question,” says Darrell Schlom, a materials scientist from Cornell University in Ithaca, New York. So, many researchers are now trying to integrate oxide heterostructures into silicon wafers. “The plan is not only to integrate oxides with silicon electronics, but even more importantly to take advantage of the processing infrastructure of silicon technology,” says Schlom.

This is an arduous task, not least because there are substantial differences in crystal structure between most oxides and silicon. And worse, oxide thin films are grown by condensing a high-temperature vapour that includes oxygen — which can turn silicon into silicon dioxide at the slightest contact. This can be avoided only by carefully adjusting growth temperatures and supplying just the right amount of oxygen at precisely the right time. Still, progress has been made and the quality of oxide films on silicon has been improving steadily<sup>11</sup>. “Even advanced oxide films such as  $\text{LaAlO}_3/\text{SrTiO}_3$  heterostructures have now been grown on silicon,” says Mannhart.

### Not so crazy

Particularly promising in this regard is zinc oxide ( $\text{ZnO}$ ), which is itself a semiconductor with a wide range of potential applications. “In  $\text{ZnO}$ , electrons can travel up to a micrometre without scattering,” says Kawasaki. Kawasaki and his colleagues have even observed the quantum Hall effect in  $\text{ZnO}$  — a first for an oxide<sup>12</sup>. The presence of such quantum phenomena suggests the use of  $\text{ZnO}$  for ‘spintronics’ applications, which promise ultrahigh-density storage and ultrafast processing of information using the electron’s tiny magnetic moment, or spin.

This isn’t the end of the possible uses of oxides. “This might be a very crazy idea, but we are wondering whether these heterostructures can be applied to new types of solar cells,” says



J. MANNHART

Electronic circuits made from thin oxide layers are only starting to tap the potential of oxides.

Tokura. Solar cells are currently made of semiconductors, he explains, and function through the absorption of light with energies larger than a certain threshold known as the band gap. If the light has an energy much larger than this band gap, the excess is wasted into heat. But if electrons are confined, for example, in semiconductor nanoparticles, they begin to interact strongly with each other, which amplifies a process in which the excess energy is not wasted but rather used to excite multiple electrons. The entire process becomes more efficient.

In complex oxides, with their strong electron correlations, such an amplification could be very strong, says Tokura. Indeed, researchers already know of certain oxides in which light can excite so many electrons that the material becomes metallic. But that would still leave the problem of extracting these electrons from the oxide to put their energy to use. Even here,

oxide thin-film structures may offer a solution. The layers are generally very thin, which means that electrons generated in one film could easily be extracted to an adjacent layer. “If we can make this work, it would be really exciting,” says

Kawasaki, who is investigating this idea with Tokura.

Parkin has an even more ambitious idea. He is looking for layered oxide systems in which superconductivity sets in at unprecedentedly high temperatures. “Room temperature is, of course, the ultimate goal,” Parkin says. “In my mind this is entirely feasible.” He thinks that such superconductivity might be found at interfaces similar to  $\text{LaAlO}_3/\text{SrTiO}_3$ , and might also involve the use of oxide compounds that do not normally exist in nature and can only be stabilized as thin films.

After more than 20 years of research into

oxide thin films, efforts are bearing fruit. Progress is becoming fast-paced. Thin-film-growth technology has been adapted for oxide compounds, suitable substrates have been developed and complex heterostructures are being studied for new functionality. “Although what we have achieved as a community is still at the very early stages, we now know a lot more about the basic rules of engagement,” says Hwang.

At the same time, Hwang sounds a note of caution. “Now the hard questions come,” he says. Even seemingly mundane issues such as sample quality need to be tackled. “Oxide heterostructures are still loaded with defects. Understanding how to control these is key to taking oxide heterostructures from scientific curiosity — their current position in various scientific sandboxes — to real technologies,” says Schlom.

Nevertheless, the achievements so far are strong testament to the fact that researchers in the field have begun to predict and control the phenomena that can exist in oxide heterostructures. Whether as new electronic compounds, as sensors, as memory devices, as solar cells or simply for their exciting science, oxide heterostructures are here to stay. The journey has merely begun. ■

Joerg Heber is a senior editor of *Nature Materials*.

- Ohtomo, A. & Hwang, H. Y. *Nature* **427**, 423–426 (2004).
- Jin, S. *et al. Science* **264**, 413–415 (1994).
- Ohtomo, A., Muller, D. A., Grazul, J. L. & Hwang, H. Y. *Nature* **419**, 378–380 (2002).
- Nakagawa, N., Hwang, H. Y. & Muller, D. A. *Nature Mater.* **5**, 204–209 (2006).
- Reyren, N. *et al. Science* **317**, 1196–1199 (2007).
- Ramesh, R. & Spaldin, N. A. *Nature Mater.* **6**, 21–29 (2007).
- Kanki, T., Tanaka, H. & Kawai, T. *Appl. Phys. Lett.* **89**, 242506 (2006).
- Salvador, P. A., Haghir-Gosnet, A.-M., Mercier, B., Hervieu, M. & Raveau, B. *Appl. Phys. Lett.* **75**, 2638–2640 (1999).
- Cavaglia, A. D. *et al. Nature* **456**, 624–627 (2008).
- Kawasaki, M. *et al. Science* **266**, 1540–1542 (1994).
- Warusawithana, M. P. *et al. Science* **324**, 367–370 (2009).
- Tsukazaki, A. *et al. Science* **315**, 1388–1391 (2007).

**“This might be a crazy idea, but maybe these heterostructures can be applied to solar cells.”**

— Yoshinori Tokura

## CORRESPONDENCE

These letters respond to the Essay 'Do nations go to war over water?' by Wendy Barnaby  
(*Nature* **458**, 282–283; 2009).

## The potential for water conflict is on the increase

SIR — Wendy Barnaby dispels the myth of future water wars, arguing that countries "solve their water shortages through trade and international agreements". But water has often been the object, target or weapon of military or terrorist action, including in former conflicts between Iraq and Iran, Syria and Iraq, Israel and Syria, and South Africa and Lesotho (see Peter H. Gleick's water conflict chronology at <http://tinyurl.com/ce776f>). These range from small-scale demonstrations and border skirmishes to the destruction of irrigation and water-supply systems and the poisoning of water supplies.

Problems can arise when humans have too little or too much water, or water that is too dirty, and countries may accuse one another of amplifying these problems. But if a pollution incident in one country kills biological life in an international river, the downstream riparian nations will collaborate to reduce damage and prevent reoccurrence; likewise, countries usually cooperate after a disastrous flood. More difficult conflicts can arise in areas of water scarcity, but nations are increasingly willing to negotiate rather than to go to war.

A handful of arid-climate countries receive nearly all their water from abroad by means of shared rivers. The water resources of Turkmenistan, Egypt, Uzbekistan, Syria and Israel depend on neighbouring countries. Several countries successfully share international rivers within the framework of river commissions, as well as lakes and aquifers through international bodies. Small-scale water disputes can lead to the development of diplomacy and help to prevent bigger water wars.

However, the potential for water conflict is on the increase, as populations in water-stressed areas continue to grow and the demand for water increases to improve living standards with better sanitation and a water-intensive diet. In arid areas, water scarcity is likely to be exacerbated by climate change (B. Bates *et al.* *Climate Change and Water* Intergovernmental Panel on Climate Change; 2008; <http://tinyurl.com/c4kd2e>). The potential for water conflict — war or no war — deserves careful attention.

**Zbigniew W. Kundzewicz** Institute for Agricultural and Forest Environment, Polish Academy of Sciences, Bukowska 19, 60-809 Poznań, Poland, and Potsdam Institute for Climate Impact Research, Telegrafenberg, 14412 Potsdam, Germany  
e-mail: [kundzewicz@yahoo.com](mailto:kundzewicz@yahoo.com)  
**Piotr Kowalczak** Institute for Agricultural and Forest Environment, Polish Academy of Sciences, Bukowska 19, 60-809 Poznań, Poland

## Increasing inequality is already making shortages worse

SIR — Wendy Barnaby is hasty in her conclusion that water shortages can and will be resolved through international trade and economic development. She predicts that nations will be able to adapt to water shortages as they develop, by switching to industries other than agriculture and by importing their food from elsewhere. But this assumes both that developing nations will be able to afford food and that food itself is not also a limited resource being threatened by global warming.

Barnaby's implication that poorer nations will become wealthier in the coming decades is at odds with the global reality of an increasing gap between rich and poor (*Report on the World*

*Social Situation* United Nations Department of Economic and Social Affairs; 2005) and with the repeated failure of such development plans (J. Ferguson *Global Shadows* Duke Univ. Press; 2006). Evidence of the causal relationship between inequality and the failure of economic development (W. Easterly *J. Dev. Econ.* **84**, 755–776; 2007) makes it ever harder to imagine that development is going to become successful in the near future.

In addition to the increasing inequality of nations, the gap between rich and poor citizens is widening in countries around the world. This means that the belief in cooperation as the "dominant response to shared water resources" needs to be further questioned by analysing how water is distributed within nations. For instance, 'water management devices' (meters) are being installed in poor people's homes in Cape Town, South Africa; in Chile, water privatization has allowed mining companies to monopolize scarce water supplies, forcing people to abandon their home towns as water dries up.

Access to water is already a serious issue for people in many parts of the world and, given recent United Nations estimates, the situation is not likely to improve. In light of this, we believe that embracing international trade and development as easy or inevitable solutions is naive. The reality of global inequality, power and poverty — and their impact on access to water — must be taken seriously in any plan for addressing future water shortages.

**Thomas H. Meek** Department of Biology, University of California, 900 University Avenue, Riverside, California 92521, USA  
e-mail: [tmeek001@ucr.edu](mailto:tmeek001@ucr.edu)  
**Laura A. Meek** Women's Studies Department, The George Washington University, 837 22nd Street NW, Washington DC 20052, USA  
e-mail: [Laura6@gwu.edu](mailto:Laura6@gwu.edu)

## Water is a source of cooperation rather than war

SIR — Water, like any scarce common resource, creates competition that can lead to conflict, but, as Wendy Barnaby concludes in her Essay, wars are not waged over water. Her message is borne out by my own experience as senior adviser to the Israeli Water Authority.

The Jordan–Israel Peace Treaty, signed in October 1994, includes a chapter on water that has been in operation for more than 14 years to the satisfaction of both parties. Jordan's chief water negotiator, Munther J. Haddadin, in a review of the history that led up to the treaty, notes that "The claim that the June war of 1967 was a 'water war' is not substantiated" (M. J. Haddadin *Diplomacy on the Jordan* Kluwer; 2001).

In 1995 Israel signed the Oslo II Interim Agreement with the Palestinian Authority, in which Article 40 ('Water and sewage') was intended to serve for a period of five years while a permanent agreement was drawn up. Although this has not yet happened, both sides adhere to the interim agreement as a basis for coordinated management of their water resources. Despite the difficult security situation, the Joint Water Committee set up by the agreement has met and communicated regularly, and Israel has increased the water supply to the Palestinian Water Authority over and above that stipulated in the agreement, in response to Palestinian needs and requests.

Water, by its very nature, has a proven potential for engendering cooperation between nations rather than being a cause for war.  
**Uri Shamir, Stephen Grand, Nancy Grand** Water Research Institute, Technion – Israel Institute of Technology, Haifa 32000, Israel  
e-mail: [shamir@technion.ac.il](mailto:shamir@technion.ac.il)



## BOOKS &amp; ARTS

## Dissecting The Two Cultures

Fifty years ago today, Charles Percy Snow argued in an influential lecture that the failure of science and the humanities to converse, and the lack of scientists in positions of power, was disastrous for society. In the first of three essays marking this anniversary, **Martin Kemp** contends that the real enemy of understanding is not these 'Two Cultures' but specialization in all disciplines.

"The Two Cultures" is a phrase — like "the corridors of power" — that has seeped into common usage. Divorced from their original context, such phrases tend to become a form of negligent shorthand that allows us to avoid precise thinking. Both were coined by the same author, Charles Percy Snow — one-time physicist, prolific novelist and political climber.

'The Two Cultures' was the title of Snow's hugely influential Rede Lecture at the University of Cambridge, UK, on 7 May 1959. One culture was science; the other was the humanities, as represented by "literary intellectuals". Snow decried what he saw as the total inability of highly educated people to cross a deep rift of mutual incomprehension.

Snow's cultural diagnosis is encapsulated in his famous challenge: "Once or twice I have been provoked and have asked the company [of 'intellectuals'] how many of them could describe the Second Law of Thermodynamics. The response was cold: it was also negative. Yet I was asking something which is about the scientific equivalent of: 'Have you read a work of Shakespeare's?'"

Snow's early career as a research scientist had been aborted in the mid-1930s, when he became disillusioned by having to acknowledge that some of his experimental work on vitamin A — published in *Nature* with Philip Bowden — did not stand up to scientific scrutiny. He subsequently flourished as a novelist, most notably with his 'Strangers and Brothers' series, centered incestuously in the hermetic hothouse of Cambridge academic politics.

He also climbed the ladder of official posts, rising to become parliamentary secretary to the Minister of Technology (1964–1966) in the House of Lords as Baron Snow. He was a major shaper of Prime Minister Harold Wilson's 1963 vision of the "white

heat of the technological revolution", and commanded a wide audience both nationally and internationally.

Snow saw applied science as holding the key to a humane future, in terms of a rational understanding of nature but also as the only force that could tackle the problems of well-being in developed and developing countries. Yet 'Luddites' from the humanities still prevailed in the 'corridors of power' — as Snow titled his 1964 novel.

### Fuelling the fire

In 1960, as a student newly arrived at the University of Cambridge, I inadvertently encountered the person who was to reignite the controversy sparked by Snow. I saw a haggard figure, shuffling across the lawns at Downing College. He was draped in an elderly coat intended for a more ample frame. His leathery neck emerged from a shirt with no tie. Not knowing who he was, I gave him a wide berth, wary of being asked for money.

It transpired that this was the legendary don of English literature and fiery literary critic, Frank R. Leavis. In 1962 Leavis subjected Snow and 'The Two Cultures' to a stinging assault, described not unfairly by philosopher Simon Critchley as "a vicious *ad hominem* attack". Leavis delivered his criticism in the Richmond Lecture that commemorated the last of his 30 years of teaching at Downing College.

Despite his international reputation, Leavis remained, and relished remaining, an outsider in official university circles. Snow, by contrast, had become a heavyweight of the establishment. The

cover of the slim volume of Snow's lecture in the university book shop portrayed a well-nourished bulldog of a man in contemplative mode, with dark jacket and neat tie. An obvious insider, in contrast to Leavis the outsider.

Leavis despised Snow's literary works: "as a novelist he doesn't exist; he doesn't begin to exist. He can't be said to know what a novel is." Leavis also dismissed Snow's authority as a cultural guru, regarding him as a mindless sign of the times: "he is a portent in that, being in himself negligible, he has become for a vast public on both sides of the Atlantic a master-mind and a sage ... It is ridiculous to credit him with any capacity for serious thinking about the problems on which he offers to advise the world."

Leavis acclaimed great literature as the true guardian of human values: "the judgments the literary critic is concerned with are judgments about life. What the critical discipline is concerned with is relevance and precision in making and developing them." He sided with Blaise Pascal, the French seventeenth-

century mathematician and theologian, who declared in his *Pensées* that "physical science will not console me for the ignorance of morality in the time of affliction. But the science of ethics will always console me for the ignorance of the physical sciences."

For Leavis, science — and the technological society it was spawning — was devoid of humane values. He insisted on the need for other kinds of concern, "entailing forethought, action and provision about the human future". To speak of human well-being only "in terms of productivity, material standards of living, hygienic and technological progress" was morally bankrupt. Leavis was witnessing with horror what he saw as the beginning of a takeover by dreaded technocrats, the apocalyptic results of which had been portrayed by George Orwell in his 1949 novel *Nineteen Eighty-Four*.

### False divide

Viewed historically, Snow's way of setting up the debate about the two cultures was founded on a false comparison between knowledge of Shakespeare and thermodynamics. The roots

**"In almost all countries, a gulf of understanding has opened up by the time students enter university."**



Science stalwart: Charles Percy Snow.



of this mistaken comparison were laid when knowledge in all forms of learning started to become specialized and professionalized, reaching an apogee when disciplines were institutionalized in the nineteenth century. The establishment of societies was not limited to the sciences and technologies. We can set, for instance, the founding of the Royal College of Veterinary Surgeons in 1844 beside that of the Government School of Design (later to become the Royal College of Art) in 1837.

Since then, the general aspects of high culture have continued to engage professionals in the sciences and humanities to similar degrees. A 2006 study from University College London showed that scientists are only a little less likely to watch a Shakespeare play than their counterparts in the humanities. Specialized research in the humanities is another matter. All

academic subjects have become 'laboratory' pursuits with respect to their specialized techniques and vocabularies. Snow's poser about the second law of thermodynamics would be better matched against a narrower question in literary studies, such as asking what is meant by deconstruction as practised by the philosopher Jacques Derrida.

On re-reading the Rede  
and Richmond lectures today,

I am struck by how they are very much of their time. It is difficult to disentangle the personal animosity, the citing of anecdotal experiences and the academic politics from the real issues.

Perhaps the best statement of what was and remains at stake came in Snow's later essay "The Two Cultures: A Second Look," first published in *The Times Literary Supplement* in 1963: "Persons educated with the greatest intensity we know can no longer communicate with each other on the plane of their major intellectual concern. This is serious for our creative, intellectual and, above all, our normal life. It is leading us to interpret the past wrongly, to misjudge the present, and to deny our hopes of the future. It is making it difficult or impossible for us to take good action."

But we should also recall, above all in the light of current financial and ecological crises, Leavis's insistence on the inadequacy of defining human 'progress' in terms of the implementation of technologies that have been seen as delivering endless economic growth.

The issue does not involve two monolithic 'cultures' of science and humanities. It is about the narrow specialization of all disciplines and wider understanding. I wonder how many biologists could answer Snow's test question, especially in the light of modern physics.

I suspect that most scholars in the humanities would fare little better with the Derrida test.

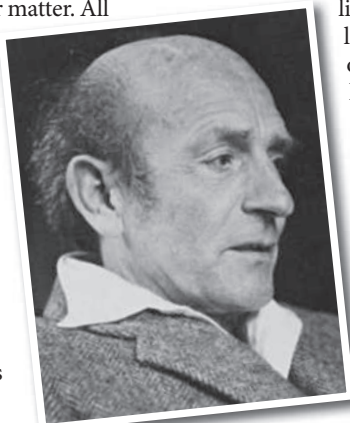
The problem is educational. There is certainly a division between ‘sciences’ and ‘humanities’, but the categories are too general to be useful in formulating any plan of action. It is the perceived need for intense specialization of any kind — in history or physics, in languages or biology — that needs to be tackled. Levels of early specialization vary across the world, but in almost all countries, a gulf of understanding has opened up by the time students enter university.

What is needed is an education that inculcates a broad mutual understanding of the nature of the various fields of research, so that we might recognize where their special competence and limitations lie. To paraphrase Christ from the Bible, it is a case of 'render unto science the things that are the sciences' and 'render unto humanities the things that are the humanities'. It is equally important not to render more to each than is warranted. The trick is to do this in the public arena, using well-informed judgement over what belongs and does not belong to each.

Snow's concern about the rift between science and the humanities is real and urgent. But so are Leavis's questions about the terms on which we can arrive at a humane definition of progress. ■

**Martin Kemp** is emeritus professor in history of art at the University of Oxford, Oxford, UK.

**See Editorial, page 10.**



**Literary legend: Frank R. Leavis.**



# Science's new battle lines

In the second of three essays on the 'Two Cultures', **Georgina Ferry** detects that today's division lies between optimists and pessimists rather than between scientific and literary intellectuals.

Optimism was not in vogue in literary circles in 1959, at the height of the cold war. Indeed, it had not enjoyed much currency throughout the twentieth century. With the decline of religion, the rise of Freudian psychology and the social and political consequences of industrialization, writers turned inward and found that the only remaining certainty was death.

On the charge sheet that physicist and novelist Charles Percy Snow drew up against such "literary intellectuals" in his Rede Lecture of that year at the University of Cambridge, UK, was the belief among writers that scientists are "shallowly optimistic, unaware of man's condition". Snow robustly countered that scientists were perfectly aware of the tragedy of the individual, but that as social beings they felt a compulsion to act to improve the lot of others. "There is plenty in our condition which is not fate, and against which we are less than human unless we do struggle," he said. He was an advocate of science — but also of hope.

Snow believed that the cultural divide he described was a recent phenomenon. It did not exist before the mid-nineteenth century, as is obvious in Richard Holmes's brilliant panorama of the cultural impact of science in the Romantic period, *The Age of Wonder* (see *Nature* 457, 31–32; 2009). Holmes places at the centre of his narrative the personal friendship between chemist Humphry Davy and poet Samuel Taylor Coleridge. The young Davy's infectious curiosity and enthusiasm delighted the poet. For his own part, Davy saw no conflict between working as a scientist and writing verse himself. The two shared experiences such as the inhalation of nitrous oxide gas as part of Davy's medical research, and foresaw the great boon that painless surgery would represent. In a letter sent to Davy on the first day of the nineteenth century, Coleridge wrote that as science was "being necessarily performed with the passion of Hope, it was poetical".

## An unsteady alloy

Today, by contrast, some cultural commentators reserve a particular kind of literary scorn for the sciences' claim of human betterment. The influential philosopher John Gray,



Writers Ian McEwan (above) and Philip Pullman (right) are optimistic about science's capacity to solve global problems.

formerly professor of European thought at the London School of Economics, is a prominent critic of Enlightenment thinking and takes a dourly pessimistic view of the capacity of humanity to behave in a selfless fashion. "The idea of progress is detrimental to the life of the spirit," he declares in his essay 'Agenda for Green Conservatism', reissued in his 2009 collection *Gray's Anatomy*.

Similarly, Martin Amis, author of novels including *Time's Arrow* and *The Information*, and short story collections such as *Einstein's Monsters*, has argued that far from being progressive, scientific advance has led to a steady demotion or "disappointment" of humanity. In a discussion on literature and science that he hosted in 2008 at the University of Manchester, UK, Amis explained how discoveries such as heliocentricity and evolution have knocked us off our Graeco-Judaean-Christian pedestal.

Expanding on the inevitability of humanity's use, or misuse, of science for baleful purposes, such as racially based eugenics, he concluded that "Human beings and science are an unsteady alloy". Asked what he meant, Amis answered: "Science means knowledge, knowledge is power, power corrupts. And that is not a metaphor — it is something the



D. LEVENSON/GETTY; CAPE FAREWELL

intelligent among the political class have always understood."

Pessimists such as Gray and Amis argue that advances in knowledge and technology do nothing for human spiritual development. Most contemporary scientists would agree. The optimism of science is twofold: that its methods might reveal, one tiny pixel at a time, more of the wonder of the natural world; and that this knowledge might be applied to solve practical human problems. There is abundant evidence in both cases that this optimism is justified; but to ask that science should do more is fundamentally to misunderstand the nature of the enterprise.

As in the time of Davy and Coleridge, the fault line today lies between optimists and pessimists, rather than between science and literature. For example, astrophysicist Martin Rees, president of the UK Royal Society, has published one of the bleakest outlooks for

**"Despair may or may not make good literature. Science, by contrast, is impossible without hope."**

humanity of recent years. Entitled *Our Final Century* (Heinemann, 2003; the US title is the even more melodramatic *Our Final Hour*), Rees's book gives human civilization no more than a 50/50 chance of surviving until 2100. What he fears is not climate change but bioterrorism, deploying the very technologies that were developed as the solutions to many medical and industrial problems.

At the same time, literary figures are beginning to line up with the optimists. Philip Pullman is the author of the internationally successful trilogy *His Dark Materials* and a regular participant in discussions on climate change. An atheist who has nevertheless read his Bible, he robustly backs St Paul's imperative to be hopeful. "Hope is the name not of a temperament, or of an emotion, but of a virtue," he says. "A virtue is something that you have to work at, something you have to do. And we can try to think and act as if it's possible to survive and to make things better, because hope is a great energizer, a comforter, an inspirer."

Novelist Ian McEwan, author of *Atonement*, is intrigued by the way science expands the possibilities of human action. The protagonist of his novel *Saturday*, Henry Perowne,

is a brain surgeon. The plot, which McEwan characteristically winds up to a pitch of almost unbearable suspense, turns on Perowne's training in neurological diagnosis and includes a lovingly detailed account of a neurosurgical operation. McEwan has maintained an informed interest in science for at least a decade, and what attracts him

is its optimism. "You can't be curious and depressed," he told an interviewer (*The Independent*, 6 April 2007). "Curiosity is itself a sure stake in life."

McEwan participated in the Cape Farewell project of 2005, in which artists and writers joined environmental scientists on an expedition to Spitsbergen in the Arctic. Ever since, as he recently revealed in *The New Yorker*, he has been working on a novel that will have a physicist as its central character, and climate change as a "background hum". The physicist is not heroic: he is physically unprepossessing, a glutton and a philanderer, yet for all his human flaws he is trying to do something good.

### Hope for the future

Snow's 1959 lecture is often seen as the first serve in a slanging match between the arts and the sciences that has been replayed fit-

fully ever since. His intention, however, was to focus on the necessity of action to alleviate the problems of the developing world, something that concerned him much more than the cold war. Today those problems are as acute as ever, and climate change will only exacerbate them. Few would now argue that science alone has all the answers, or deny that technological development has presented new problems of its own. We are left with two choices. We can either regret the massive social and global changes that have accompanied the shift to a largely technologically driven society, and predict humanity's decline; or we can use the skills we have — including science but also politics, art and literature — to try to mitigate the worst evils.

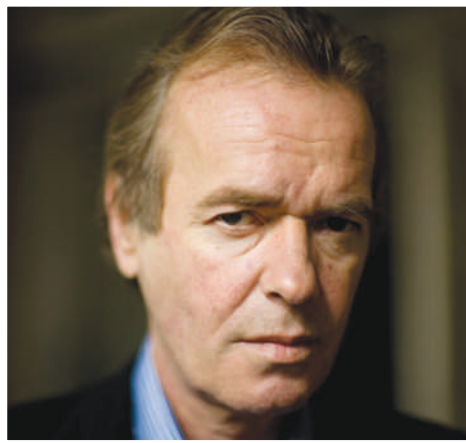
Despair may or may not make good literature. Science, by contrast, is impossible without hope. But it is not a shallow, Panglossian kind of hope. It is simply the energy to grapple with a problem and think about how it might be solved. We might well be adrift in a meaningless Universe, but we might as well try to make the best of it. ■

**Georgina Ferry** is a writer based in Oxford, UK. Her latest book is *Max Perutz and the Secret of Life*.

See Editorial, page 10.



ILLUSTRATION BY SUZANNE BARRETT



Astronomer Martin Rees (left) fears that the human race may not survive the next century; author Martin Amis (right) worries that science may be misused.



# Snow's portrait of science in politics

Charles Percy Snow ignited controversy around science and policy-making in a series of lectures at Harvard University a year after his 'Two Cultures' debate. Below we reproduce an extract from the resulting book, *Science and Government*. It gives a remarkable insight into how science feeds into political decision-making.

Snow was one of Britain's leading science civil servants. He capitalized on his experience at the heart of the government machinery to analyse the role of science and scientists in the United Kingdom during the Second World War. He wanted to disentangle how political decisions were made during the war and, importantly, how scientific advice was used to make them.

In the extract, Snow gives a colourful account of the decision to invest in the development of radar — an unproven technology in the mid-1930s that would prove to be key to the success of the allied war effort.

In the years leading up to the war, the British government did not know how to handle the rise of Adolf Hitler and his military build-up

in Germany. Snow homes in on two men, both scientists. One is the chemist Henry Tizard — rector of Imperial College, London, and chairman of an influential committee that advised the Labour government to invest in radar technology. The other is Tizard's one-time friend and later rival, Frederick Lindemann, a close adviser to Winston Churchill, who was then part of the Conservative opposition.

Such an account could have been dry and factual, but this is not what we get from Snow's pen. Instead, his readers are treated to a page-turner. Snow the novelist is not afraid to tell us what his characters might be thinking. He is indiscreet about their personal lives, social connections and professional backgrounds. He guesses their motives and provides his own verdict on their judgements.

Snow was attacked for not being detached enough in his analysis. Yet the book's accessibility ensured it was widely read. Five decades later, *Science and Government* remains a rare inside view of the relationship between science and political authority. **Joanne Baker**

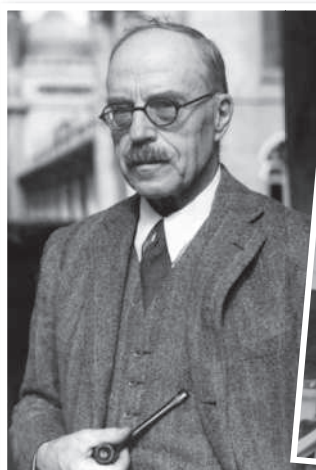
**Extract from**  
*Science and Government*  
by C. P. Snow  
Harvard University Press: 1961.

In 1934 both Tizard and Lindemann were nearly fifty. Of the two, Tizard had been by a long way the more successful, though even he, judged by the standard he set himself, had not lived up to his promise. He was a trusted man of affairs, he had been knighted, he was head of a university institution, but in his own eyes he had not done much.

As for Lindemann, he had done much less. The professional physicists did not take him seriously as a scientist, and dismissed him as a cranky society pet. Scientifically his name was worth little. He was the intimate friend of [Churchill] a politician whose name was scarcely worth as much.

Then, quite suddenly, Tizard was given the chance for which he was made. England was strategically in a desperately vulnerable position, for reasons — the tiny size of the country, the density of the population — which apply more harshly today. In 1934 [Stanley] Baldwin was the main figure in the government, and it was only two years since he had said lugubriously: "The bomber will always get through."

In public, rebellious politicians like Churchill were attacking the whole of the government's defence policy. In secret, the government scientists, the military staffs, the high officials,



**Frederick Lindemann (right) vehemently disagreed with Henry Tizard (left) over the development of radar as Britain's method of air defence.**

were beating round for some sort of defence. There was nothing accidental about this. It was predictable that England, more vulnerable to air attack than any major country, would spend more effort trying to keep bombers off. But there was something accidental and unpredictable in Tizard being given his head.

The Air Ministry, under the influence of their scientific adviser, H. E. Wimperis, himself prodded by a bright young government scientist called A. P. Rowe,<sup>1</sup> set up a Committee for the Scientific Study of Air Defence. Its terms of reference were as flat as usual: "To consider how far advances in scientific and technical knowledge can be used to strengthen the present methods of defence against hostile aircraft." The committee was nothing very

important to start with. No one took much notice when its membership was announced. There may have been slight curiosity about the appointment, which was entirely due to Wimperis,<sup>2</sup> of Tizard as chairman. The appointment would not and could not have happened, though, if Tizard had not been so well connected in official life.

Well, that committee was called the Tizard Committee almost from its first meeting. It is slightly touching that in his diary Tizard, who could not use that title, never seems to have been quite certain what its official title really was.

From the first meeting on January 28th, 1935 he gripped the problems. This was the job for which

he was born. Quite soon, by the summer of that year, small ripples of confidence oozed under the secret doors and penetrated Whitehall, almost the only ripples of confidence that touched the official world during those years. Tizard insisted on a very small committee which he chose himself. Wimperis had to be there, Rowe was brought in as secretary, but at the beginning there were only two members of independent standing, A. V. Hill and P. M. S. Blackett. Both of these were eminent scientists, of a quite different order of accomplishment from Tizard or Lindemann. Hill was one of the most distinguished physiologists in the world and had won a Nobel prize in 1922. Blackett, who was only thirty-seven at this time, was one of [Ernest] Rutherford's

POPPER/GETTY; CENTRAL PRESS/HULTON ARCHIVE/GETTY

most brilliant pupils, and later himself won a Nobel prize.<sup>3</sup> ...

The committee made up its mind about [radar] before the device really existed. Watson Watt, who was the pioneer of radar in England, ... had done some preliminary experiments. This device might, not certainly but possibly, work in real war in three or four years. Nothing else possibly could. Tizard, Hill, Blackett had faith in their own reasoning. Without fuss, and without backward glances, the choice was made. That was only a resolution on paper, and they had to make it actual.

The administrative mechanism by which this was done is itself interesting. In form the Air Minister, Lord Swinton,<sup>4</sup> arranged for a new high-level committee ... Over this new body he himself presided, and on to it was brought the government's chief military critic, Winston Churchill. In fact, however, one has got to imagine a great deal of that apparently casual to-ing and fro-ing by which high English business gets done. As soon as the Tizard committee thought there was something in radar, one can take it that Tizard would lunch with Hankey<sup>5</sup> at the Athenaeum; Hankey, the secretary of the Cabinet, would find it convenient to have a cup of tea with Swinton and Baldwin. If the Establishment had not trusted Tizard as one of their own, there might have been a waste of months or years. In fact, everything went through with the smoothness, the lack of friction, and the effortless speed which can only happen in England when the Establishment is behind one. Within a very short time the Tizard Committee were asking for millions of pounds, and getting it without a blink of an eye. Two successive secretaries of the Cabinet, Hankey and Bridges,<sup>6</sup> did much more than their official duties in pushing the project through.

The second active job was, in particular, to persuade the serving officers of the Air Staff that radar was their one hope and, in general, to make scientists and military people understand each other. Here again this might have been impossible. In fact, with the exception of those concerned with bombing policy, the senior officers were ready to be convinced as soon as Tizard started to talk.<sup>7</sup> They often thought of putting him in uniform: but that would have defeated his whole

virtue as an interpreter between the two sides. "I utterly refuse to wear a busby," he used to say. Fairly soon he had not only got radar stations in principle accepted and hoped for, but also succeeded, with the help of Blackett's exceptional drive and

important decisions were in effect taken. By the end of 1936 most of those decisions were translated into action. It was one of the most effective small committees in history. But before it clinched its choices, there was a most picturesque row.

The committee had been set up, as we saw, from inside the Air Ministry. One of the reasons was, no doubt, to forestall criticism from outside, which came most loudly and effectively from Churchill. In 1934 he had publicly challenged the government's underestimate of the size of Hitler's air force. His figures, which had been produced by Lindemann, were much nearer the truth than the government's. Thus, simultaneously, there were going on the secret deliberations and discussions of the Tizard Committee, and an acrimonious military argument in full light in the House of Commons and the press, with Churchill the antigovernment spokesman.

It is one of the classical cases of "closed" politics coexisting with "open" politics. Passing from one to the other, an observer would not have known that he was dealing with the same set of facts. By the middle of 1935 Baldwin, who had just in form as well as fact become Prime Minister, wanted to reduce the temperature of the "open" military argument. He used the orthodox manoeuvre of asking Churchill in. Not into the Cabinet: the personal rifts were too deep for that, but onto the new Swinton Committee, [a] political committee ..., which was to keep a supervisory eye on air defence.

The history is very tangled at this point. No minutes have ever been published, but if I know Hankey and his colleagues at all — and I had the good luck to work under them a short time later — I have not much doubt that on the one hand they felt confident that they could give the Tizard Committee its head (Tizard sat himself on the political committee and made his requests for money to it), and that on the other hand it could not do harm, and might do good, if Churchill were given exact information of what was actually being done, rather than inexact.

Roughly that was what happened, but there were other consequences. Churchill entered the political committee, retaining the right to criticise in public and insisting that Lindemann, as his personal scientific adviser, be given a place on the Tizard Committee. Both



insight, in beginning to teach one lesson each to the scientists and the military, lessons that Tizard and Blackett went on teaching for twenty years.

The lesson to the military was that you cannot run wars on gusts of emotion. You have to think scientifically about your own operations. This was the start of operational research,<sup>8</sup> the development of which was Blackett's major personal feat in the 1939–45 war.<sup>9</sup> The lesson to the scientists was that the prerequisite of sound military advice is that the giver must convince himself that, if he were responsible for action, he would himself act so. It is a difficult lesson to learn. If it were learnt, the number of theoretical treatises on the future of war would be drastically reduced.

The committee met for the first time, as I said, in January 1935. By the end of 1935 its



these conditions were reasonable enough: but then the private war began.

Almost from the moment that Lindemann took his seat in the committee room, the meetings did not know half an hour's harmony or work undisturbed. I must say, as one with a taste for certain aspects of human behaviour, I should have dearly liked to be there. The faces themselves would have been a nice picture. Lindemann, Hill, and Blackett were all very tall men of distinguished physical presence — Blackett sculptured and handsome, Hill ruddy and English, Lindemann pallid, heavy, Central European. Blackett and Hill would be dressed casually, like academics. Tizard and Lindemann, who were both conventional in such things, would be wearing black coats and striped trousers, and both would come to the meetings in bowler hats. At the table Blackett and Hill, neither of them especially patient men nor overfond of listening to nonsense, sat with incredulity through diatribes by Lindemann, scornful, contemptuous, barely audible, directed against any decision that Tizard had made, was making, or ever would make. Tizard sat it out for some time. He could be irritable, but he had great resources of temperament, and he knew that this was too serious a time to let the irritability flash. He also knew, from the first speech that Lindemann made in committee, that the friendship of years was smashed.

There must have been hidden resentments and rancours, which we are now never likely to know and which had been latent long before this. No doubt Lindemann, who was a passionate man, with the canalised passion of the repressed, felt that he ought to have been doing Tizard's job. No doubt he felt, because no one ever had more absolute belief in his own conclusions, that he would have done Tizard's job much better, and that his specifics for air defence were the right ones, and the only right ones. No doubt he felt, with his fanatical patriotism, that Tizard and his accomplices, these Blacketts, these Hills, were a menace to the country and ought to be swept away.

It may have been — there are some who were close to these events who have told me so — that all his judgments at these meetings were due to his hatred of Tizard, which had burst out as uncontrollably as love. That is, whatever Tizard wanted and supported, Lindemann would have felt unshakably was certain to be wrong and



"But the trouble is when you get on to any kind of moral escalator, to know whether you're ever going to be able to get off." — C. P. Snow (1961).

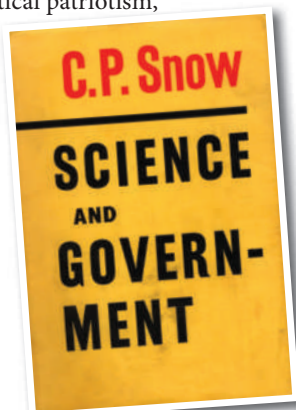
would have opposed. The other view is that Lindemann's scientific, as well as his emotional, temperament came in: it was not only hatred for Tizard, it was also his habit of getting self-blindingly attached to his own gadgety ideas that led him on. Whatever the motive was, he kept making his case to the committee in his own characteristic tone of grinding certainty. It was an unjustifiable case.

The issue in principle was very simple. Radar was not yet proved to work: but Tizard and the others, as I have said, were certain that it was the only hope. None of them was committed to any special gadget. That was not the cast of their minds. There was only a limited amount of time, of people, of resources. Therefore the first priority must be given to radar — not only to making the

equipment, but to making arrangements, well in advance even of the first tests, for its operational use. (It was in fact in the operational use of radar, rather than in the equipment, that England got a slight tactical lead.)

Lindemann would not have any of this. Radar was not proved. He demanded that it should be put much lower on the priority list and research on other devices given the highest priority. He had two pet devices of his own. One was the use of infra-red detection. This seemed wildly impracticable then, to any of the others and to anyone who heard the idea. It seems even more wildly impracticable now. The other putative device was the dropping, in front of hostile aircraft, of parachute bombs and parachute mines. Mines in various forms had a singular fascination for Lindemann.

For twelve months Lindemann ground on with his feud on the committee. He was tireless. He was ready at each meeting to begin again from the beginning. He was quite



The racy style of Snow's book meant it was widely read.

V. WEISZ/SOLO SYNDICATION/ASSOCIATED NEWSPAPERS LTD/BRITISH CARTOON ARCHIVE, UNIV. KENT, WWW.CARTOONS.AC.UK

OXFORD UNIV. PRESS



HULTON-DEUTSCH COLLECTION/CORBIS

C. P. Snow championed openness but recognized that radar would not have been developed without 'closed' decision-making.

unsoftened, quite impregnable to doubt. Only a very unusual man, and one of abnormal emotional resistance and energy, could sit with men so able and not be affected in the slightest regard.

They themselves were not affected so far as choice was concerned. Tizard went ahead with the radar decisions and let Lindemann register his disagreement. But gradually they got worn down. Neither Blackett nor Hill was phlegmatic enough to endure this monomaniac tension for ever. In July 1936,<sup>10</sup> when the committee were preparing a report, Lindemann abused Tizard in his usual form, over the invariable issue of too much priority for radar, but in terms so savage that the secretaries had to be sent out of the room.<sup>11</sup>

At that point Blackett and Hill had had enough of it. They resigned and did not try to give an emollient excuse for doing so. Whether this was done after discussion with Tizard is not clear. No discussion was really necessary. They all believed that this friction was doing too much harm. They were all experienced enough to know that, with Churchill still out of office, they could make their own terms.

Within a short time the committee was reappointed. Tizard was still chairman, Blackett and Hill were still members. Lindemann, however, was not. He was replaced by E. V. Appleton, the greatest living English expert on the propagation of radio waves. Radar itself was an application of Appleton's fundamental work. The announcement of his name meant, in the taciturn eloquence of official statements, a clear victory for radar and for Tizard. The radar stations and the radar organisation were ready, not perfect but working, in time for the Battle of Britain. This

had a major, and a perhaps decisive, effect.

This cautionary story of the first Lindemann–Tizard collision seems to me to contain a number of lessons, some of them not obvious. But there is one, at the same time so obvious and so ironic that I shall mention it now. It is simply that the results of closed politics can run precisely contrary to the results of open politics. That is an occupational feature of the way in which closed politics works and the way in which secret choices are made. Probably not more than a hundred people had any information whatever about Tizard's first radar decision; not more than twenty people took any effective part in it, and at the point of choice not more than five or six.

While that was going on, so also was violent open politics, the open politics of the thirties, the most ferocious and deeply felt open politics of my lifetime. Nearly everyone I knew of my own age who was politically committed, that is, who had decided that fascism had at all costs to be stopped, wanted Churchill brought into government. Partly for his own gifts, partly as a symbol of a country which was not going to let the Nazis win by default. We signed collective letters about Churchill; we used what influence we had, which in those years was not much. We wanted a government which would resist, the kind of government we finally got in 1940. That was the position, I think, of Blackett and most of my liberal friends. It was certainly my own. Looking back, I think we were right, and if put back in those years again I should do what I did then.

The ifs of history are not very profitable — but if Churchill had been brought back to office, if open politics had gone the way my

friends and I clamoured and implored that it should? We should, without any question, have been morally better prepared for war when it came. We should have been better prepared in the amount of war material. But, studying the story I have just told, I find it hard to resist the possibility that, in some essential technical respects, we might have been worse prepared. If Churchill had come into office, Lindemann would have come with him, as happened later. It is then very hard to imagine Lindemann not getting charge of the Tizard Committee. As I have said, I take a pretty Tolstoyan view of history in the large. In a broad sense I cannot easily accept that these small personal accidents could affect major destinies.

And yet... without getting the radar in time we should not have stood a good chance in the war that finally arrived. With Lindemann instead of Tizard, it seems at least likely that different technical choices would have been made. If that had been so, I still cannot for the life of me see how the radar system would have been ready in time.

These retrospective fears are not profitable. But I do not know of a clearer case where open and closed politics appear to tell such different stories and point to such different fates. ■

Reprinted by permission of the publisher from *Science and Government* by C. P. Snow, pp. 23–38, Harvard University Press, Cambridge, Mass. © 1960, 1961 by the President and Fellows of Harvard College. © renewed by Charles Percy Snow.

#### Endnotes

<sup>1</sup>Rowe played an important part, easy to underestimate because the whole of it was secret, in the scientific war, 1935–45. He is best known as the superintendent of the Telecommunication Research Establishment, the most brilliant and successful of the English wartime research establishments.

<sup>2</sup>It is worth noticing that Wimperis, who was a peace-loving, sweet-natured man, ill-at-ease among violent disputes, both got the committee going and selected Tizard.

<sup>3</sup>In 1948.

<sup>4</sup>Lord Swinton's part in these preparations, like Rowe's, though for different reasons, has been constantly underestimated.

<sup>5</sup>At this time Sir Maurice, later Lord, Hankey. One of the great invisible influences in English affairs, particularly military affairs, for a generation. His part has not yet been properly described.

<sup>6</sup>Later head of the Civil Service and now Lord Bridges.

<sup>7</sup>Cf. P. M. S. Blackett, "Tizard and the Science of War," *Nature* **185**, 647–653 (1960).

<sup>8</sup>"Operations research" in the United States. But the English started it, and I much prefer the English name. In the 1914–18 war, A. V. Hill's scientists were testing anti-aircraft gunnery and were carrying out what we should later have called operational research.

<sup>9</sup>P. M. S. Blackett, "Operational Research," *Brassey's Annual* (1953) 88–106.

<sup>10</sup>Not 1937 as stated in [W. S.] Churchill [*The Second World War* Vol. 1 (Cassell, 1948)], p. 120. There are other inaccuracies in the chapter ("Problems of Sea and Air, 1935–1939", pp. 115–128).

<sup>11</sup>This is Blackett's account. Rowe is inclined to think, without being certain, that this critical quarrel took place before a meeting. It may easily have happened that, since a row was expected, the secretaries were told not to come in at the beginning.

See Editorial, page 10.



## NEWS &amp; VIEWS

## PALAEOANTHROPOLOGY

# *Homo floresiensis* from head to toe

Daniel E. Lieberman

**Fossils of tiny ancient humans, found on the island of Flores, have provoked much debate and speculation. Evidence that they are a real species comes from analyses of the foot and also — more surprisingly — of dwarf hippos.**

Good science requires a healthy dose of tempered scepticism — at its heart, the process involves trying to reject proposed hypotheses. So it was understandable that the announcement<sup>1,2</sup> in 2004 of the discovery of a species of dwarfed hominin, *Homo floresiensis*, from the island of Flores, Indonesia, stimulated a range of opinions, many of them sceptical, that the fossils constituted a new species and were not the consequence of some pathological condition.

Two papers in this issue, by Jungers and colleagues<sup>3</sup> and by Weston and Lister<sup>4</sup>, together with contributions to a special online issue of the *Journal of Human Evolution*, will go a long way towards addressing the sceptics' concerns. The studies provide considerable evidence — literally from head to toe — that *H. floresiensis* is a true species of hominin (that is, a species more closely related to humans than to chimpanzees and other apes). More importantly, the analyses prompt hypotheses about the human family tree that will require more fossil evidence to test.

So far, remains of *H. floresiensis* have been excavated from just a single cave, Liang Bua (Fig. 1). The fossils include a partial skeleton (LB1) plus fragments of at least half a dozen more individuals now dated to between 95,000 and 17,000 years ago<sup>5,6</sup>. These were small people, about a metre tall. Most remarkably, LB1's skull has a chimp-size brain of 417 cm<sup>3</sup> in an approximately 30-kg body. Some palaeoanthropologists hypothesized that *H. floresiensis* evolved from a non-modern species of hominin, possibly *H. erectus*, through a process called insular dwarfing that is common to islands such as Flores, in which large species undergo intense selection to become small. Archaeological data showed that *H. floresiensis* made stone tools, and hunted dwarfed elephants (*Stegodon*) and giant varanid lizards (Komodo dragons) that were also present on the island.

Such a minuscule brain in a species so recent that also made stone tools has strained credibility. Several scholars argued that the bones come from a pathological population of human pygmies suffering from some developmental syndrome that includes microcephaly<sup>7–9</sup>. All such diagnoses have proved problematic, because none accounts for the entire suite of



**Figure 1 | Fossil site — Liang Bua cave on the island of Flores.**

features evident in *H. floresiensis*, including the size and shape of the brain and cranium<sup>10–12</sup>, and the anatomy of the shoulder<sup>13</sup> and wrist<sup>14</sup>.

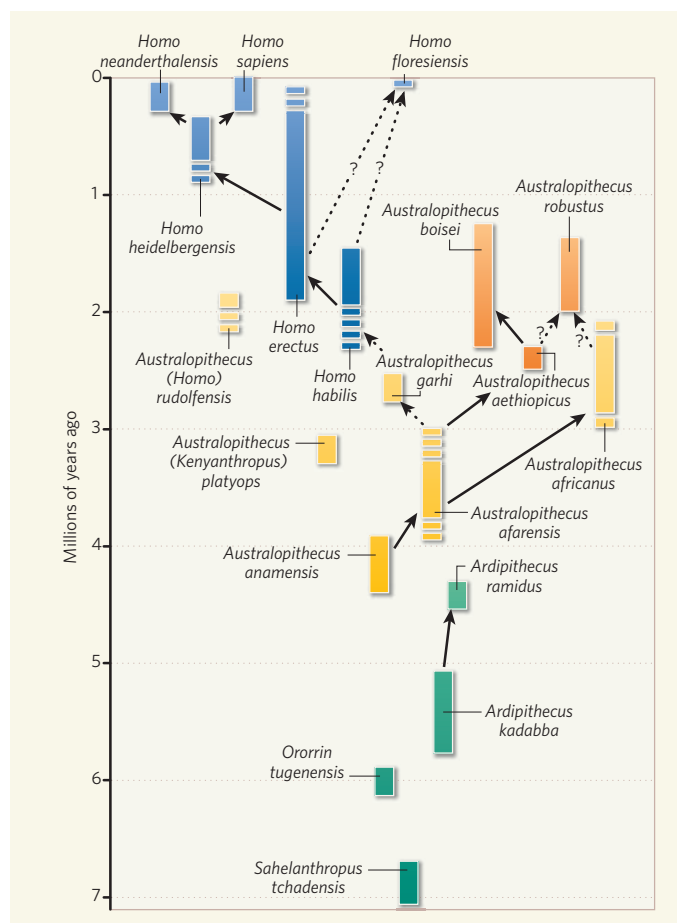
The most serious criticism, however, has been that LB1's brain is too small to be explained by known scaling relationships between brain and body size. Across species, brain mass typically scales to body mass to the power of 0.75, but among closely related species the scaling exponent is usually 0.2–0.4, and within species it is 0.25 or less<sup>15</sup>. Accordingly, if LB1 were a dwarfed human of 30 kg, then its predicted brain volume would be about 1,100 cm<sup>3</sup>; if it were a dwarfed *H. erectus* then its brain volume would be expected to be about 500–650 cm<sup>3</sup>. All in all, many scientists (myself included) have sat on the fence, waiting for more evidence about the nature and form of *H. floresiensis*.

And now we have some. One remarkable line of thinking (page 81) comes from Jungers and colleagues' description<sup>3</sup> of the species' fascinating foot. In some respects the foot is very human-like: the big toe is aligned with the

other toes; the middle of the foot apparently had a locking mechanism to help stiffen the arch after the heel lifted off the ground; and the metatarsals are typically human in several respects, including upwardly oriented joints at their ends that enable the toes to extend at the end of stance (the part of the stride when the foot is on the ground). But otherwise this is no human foot. Its approximate length, 20 cm, is much longer than one would find in any person of that stature, and instead has the proportionate length of a chimpanzee or an australopith (a genus of early hominin). Additional primitive features include long, curved and robust lateral toes; a short big toe; and a weight-bearing process on a crucial bone, the navicular, which acts like the keystone at the top of the inside of the human arch.

Together, these features suggest that the foot of *H. floresiensis* was capable of effective walking, because the middle of the foot could be stiffened when the calf muscles raised the heel off the ground. This mechanism permits the

C. TURNEY, UNIV. WOLLONGONG



**Figure 2 | The human family tree.** *Homo floresiensis* might be most closely related to early *H. erectus*, but also shows potential affinities with *H. habilis*. In either case, recognizing *H. floresiensis* as a species will require us to re-examine how we define species of the genus *Homo* and how they were related to each other. Reasonably well-known relationships are indicated by solid arrows; less secure relationships are indicated by dotted arrows. Broken vertical bars indicate uncertainties about when species evolved or went extinct.

toe flexors to push the body up and forwards at the end of stance. But the inside of LB1's arch was either weak or flat, and apparently lacked the spring-like mechanism that humans use to store and release energy during running<sup>16</sup>. In addition, the long, slightly curved toes probably posed no hindrance to walking, but would have created problematically high torques around the toe joints during running<sup>17</sup>.

It is often assumed that a human-like foot with short toes and a high arch evolved for walking. But the primitive foot of *H. floresiensis* provides a tantalizing model for a non-modern hominin foot that had evolved for effective walking before selection for endurance running occurred in human evolution<sup>16</sup>. Recently discovered footprints from Kenya indicate that a modern foot had evolved by 1.5 million years ago, presumably in *H. erectus*<sup>18</sup>. Unless the Flores fossils re-evolved a primitive foot, they must have branched off the human line before this time.

The papers in the special issue of the *Journal of Human Evolution* bolster previously published evidence that the mosaic of primitive and derived features evident in the *H. floresiensis* foot can be seen elsewhere in the skeleton. Many aspects of the anatomy, such as the scapula, are quite human-like in spite of being tiny. But there are also numerous primitive features that resemble those of either australopiths or early *Homo*. Primitive features in the upper limbs include a relatively short, very curved

clavicle; a straight humerus that lacked the normal degree of twisting between the shoulder and the elbow; and an ape-like wrist<sup>13,14,19</sup>. Primitive features in the hip and lower limbs include flared iliac blades, relatively small joints and relatively short leg bones<sup>1,20</sup>.

These features suggest that *H. floresiensis* evolved from a species that was anatomically more primitive than classic *H. erectus* from Asia. One possibility (Fig. 2) is that *H. floresiensis* evolved from *H. habilis*, whose skeleton is poorly known but is australopith-like in many respects. Another is that *H. floresiensis* descended from an earlier type of *H. erectus*, whose body may have been much less modern than we currently credit, and which perhaps deserves a separate species designation (*H. ergaster*).

But what about the head of *H. floresiensis*? LB1 has a vertical face, no snout, and most of its teeth generally resemble those of *H. erectus*. A state-of-the-art shape analysis<sup>21</sup> indicates that the LB1 skull conforms to what one predicts from a scaled-down *H. erectus* or possibly a *H. habilis*. Yet one also needs to explain how the species got such a small brain.

Here hippos come to the rescue. Weston and Lister (page 85)<sup>4</sup> analysed several species of fossil hippo that underwent insular dwarfing in Madagascar. In these species, brain mass scales to body mass to the power of 0.35 after growth has slowed in infancy, and to 0.47 when growth from birth is considered. Further, in some dwarfed species, natural selection

evidently shrank brains to volumes well below the sizes predicted by these relationships. The extra reduction presumably occurs because brain tissue is so metabolically costly that animals with relatively smaller brains can save energy when resources are scarce.

Such dwarfing is enough to account for LB1's 417-cm<sup>3</sup> brain and 30-kg body if *H. floresiensis* were a dwarfed version of the small early *H. erectus* females from Dmanisi, Georgia, that were 40 kg and had brain volumes in the range 600–650 cm<sup>3</sup> (ref. 22). Alternatively, *H. floresiensis* might be descended from *H. habilis*, whose body size was possibly just as small, about 30 kg in females. But this hypothesis, too, requires some significant brain shrinking, about 25%, because the smallest known *H. habilis* cranium (KNM-ER 1813) has a 509-cm<sup>3</sup> brain.

Overall, *H. floresiensis* presents a fascinating conundrum, and prompts some tantalizing predictions that will continue to strain credulity without more fossil evidence. First, if the species evolved from early *H. erectus*, possibly like the fossils found at Dmanisi, then this species (or group of species) was more diverse and anatomically more primitive in many respects (hands and feet for example) than previously recognized. A more audacious hypothesis is that *H. floresiensis* evolved from an even more primitive species, perhaps *H. habilis*. If so, then this species also migrated out of Africa but left no trace yet found, except on Flores. My wager is on the first possibility. But the only way to test these and other hypotheses is to find more fossils, especially in Asia. Get out your shovels!

Daniel E. Lieberman is in the Department of Anthropology, Harvard University, Cambridge, Massachusetts 02138, USA.

e-mail: danlieb@fas.harvard.edu

1. Brown, P. et al. *Nature* **431**, 1055–1061 (2004).
2. Morwood, M. J. et al. *Nature* **431**, 1087–1091 (2004).
3. Jungers, W. L. et al. *Nature* **459**, 81–84 (2009).
4. Weston, E. M. & Lister, A. M. *Nature* **459**, 85–88 (2009).
5. Morwood, M. J. et al. *Nature* **437**, 1012–1017 (2005).
6. Roberts, R. G. et al. *J. Hum. Evol.* doi:10.1016/j.jhevol.2009.01.003 (2009).
7. Jacob, T. et al. *Proc. Natl Acad. Sci. USA* **103**, 13421–13426 (2006).
8. Hershkovitz, I., Kornreich, L. & Laron, Z. *Am. J. Phys. Anthropol.* **134**, 198–208 (2007).
9. Obendorf, P. J., Oxnard, C. E. & Kefford, B. J. *Proc. Biol. Sci.* **275**, 1287–1296 (2008).
10. Falk, D. et al. *Proc. Natl Acad. Sci. USA* **104**, 2513–2518 (2007).
11. Argue, D., Donlon, D., Groves, C. & Wright, R. J. *J. Hum. Evol.* **51**, 360–374 (2006).
12. Gordon, A. D., Nevell, L. & Wood, B. *Proc. Natl Acad. Sci. USA* **105**, 4650–4655 (2008).
13. Larson, S. G. et al. *J. Hum. Evol.* **53**, 718–731 (2007).
14. Torcheri, M. W. et al. *Science* **317**, 1743–1745 (2007).
15. Martin, R. D., MacLarnon, A. M., Phillips, J. L. & Dobyms, W. B. *Anat. Rec. A* **288**, 1123–1145 (2006).
16. Bramble, D. M. & Lieberman, D. E. *Nature* **432**, 345–352 (2004).
17. Rolián, C. P. et al. *J. Exp. Biol.* **212**, 713–721 (2009).
18. Bennett, M. R. et al. *Science* **322**, 1089–1092 (2009).
19. Larson, S. G. et al. *J. Hum. Evol.* doi:10.1016/j.jhevol.2008.06.007 (2009).
20. Jungers, W. L. et al. *J. Hum. Evol.* doi:10.1016/j.jhevol.2008.08.014 (2009).
21. Baab, K. L. & McNulty, K. P. *J. Hum. Evol.* doi:10.1016/j.jhevol.2008.08.011 (2009).
22. Lordkipanidze, D. et al. *Nature* **449**, 305–310 (2007).



## ASTROPHYSICS

## Galaxy connections

James Dunlop

**A combined millimetre- and visible-light view of a forming cluster of galaxies in the young Universe adds yet another piece to the puzzle of how today's Universe of galaxies formed and evolved.**

What is the evolutionary pathway of a galaxy? And how do different galaxy populations relate to one another? On page 61 of this issue, Tamura and colleagues<sup>1</sup> get closer to answering these questions by showing that two apparently very dissimilar types of star-forming galaxy develop synchronously within a young cluster of galaxies seen 2 billion years after the Big Bang.

It is now 30 years since the advent of charge-coupled device (CCD) cameras on large optical telescopes facilitated the first meaningful searches for 'primeval' galaxies in the young Universe<sup>2</sup>. These first surveys were motivated by the natural expectation that young, rapidly star-forming galaxies would be bright at ultraviolet wavelengths owing to the presence of large numbers of very hot, short-lived massive stars. In particular, it was anticipated that these galaxies would exhibit strong ultraviolet emission lines owing to the ionization and subsequent recombination of hydrogen atoms. Because of the expansion of the Universe, the strongest such line, Lyman- $\alpha$ , would be redshifted from its original ultraviolet wavelength of 121.5 nanometres into the observable optical window at redshifts  $z > 2.5$ .

In the past decade, with the additional light grasp offered by 8-metre-class optical telescopes, this technique of finding galaxies through their Lyman- $\alpha$  emission has really come of age, and substantial numbers of young galaxies have been discovered out to redshifts  $z > 5$ . The current record holder for the most distant known object is a Lyman- $\alpha$  emitter (LAE) uncovered at  $z = 6.96$ , when the Universe was less than 1 billion years old<sup>3</sup>.

There is no doubt that LAEs fit the bill as young, star-forming galaxies. However, individually they are generally rather unspectacular objects, with low masses and typical star-formation rates of only a few solar masses per year. They certainly fall a long way short of the extreme bursts of star formation apparently required at high redshift to produce the massive, old stellar populations found in present-day elliptical galaxies<sup>4</sup>. But such super-starbursts in the young Universe have now been found, as a result of a second technological revolution in the mid-1990s — the advent of the first submillimetre- and millimetre-wave cameras<sup>5</sup>.

It turns out that in these extreme starburst galaxies, which form several hundred solar masses of stars per year, most of the ultraviolet light from the young stars is absorbed by interstellar dust grains, re-emitted in the far-infrared, and redshifted to (sub)millimetre

wavelengths. These submillimetre galaxies (SMGs) seem to be very different from even the most extreme LAEs, and understanding the origin of this difference, and the nature of any connection between these two populations, has become an important goal in current galaxy-formation research. In their study, Tamura *et al.*<sup>1</sup> shed some interesting new light on this issue, presenting the results of the first large-scale millimetre-wave survey of a known proto-cluster of LAEs at  $z = 3.1$  (ref. 6).

Even in this sort of targeted study, establishing the relationship between SMGs and LAEs is a tricky business. Optical astronomy is still far ahead of millimetre-wave astronomy in terms of image quality and information content. By nature of their selection, LAEs are discovered at a well-defined redshift, whereas the millimetre-wave image of a given patch of sky detects all very luminous dust-enshrouded galaxies along the line of sight out to extreme redshifts of  $z \sim 8$ .

This 'complete' view of the Universe offered by millimetre-wave astronomy is in fact one of its greatest strengths, and is a key driver for the forthcoming Atacama Large Millimeter/Submillimeter Array (ALMA) facility in Chile. But the image obtained by Tamura *et al.*<sup>1</sup> at a wavelength of 1.1 mm, using the AzTEC camera<sup>7</sup> mounted on the 12-metre-diameter ASTE (Atacama Submillimeter Telescope Experiment) telescope, essentially consists of 15 blobs with unknown redshifts, and rather poorly determined positions. Nevertheless, the strength of the AzTEC map is that it is 20 times bigger than previous (sub)millimetre maps of this field of the sky, and Tamura *et al.* have extracted maximum information from this new image. In particular, they present a compelling case that at least some of the SMGs lie within the  $z = 3.1$  filamentary proto-cluster traced out by the LAEs.

The evidence is quantified in terms of a statistically significant cross-correlation between the positions of the LAEs and SMGs on the sky. However, perhaps the most convincing result is the visually obvious fact that the brightest SMG uncovered in the AzTEC image lies right in the heart of the proto-cluster. At the same time, Tamura *et al.* reaffirm the distinction between LAEs and SMGs. None of the LAEs is detected in the AzTEC map; the authors were unable to achieve even a marginal detection at 1.1 mm when attempting to add the millimetre-wave fluxes extracted from the AzTEC map at the positions of all 166 LAEs. This non-detection confirms that typical LAEs contain less than



## 50 YEARS AGO

According to the present model, the solar corona consists of a gas of electrons and protons with a small mixture of heavy elements which is isothermal at a temperature of approximately a million degrees and which is in hydrostatic equilibrium in the gravitational field of the Sun. The alternative model which we propose is that the solar corona consists of trapped charged particles moving in the magnetic field of the Sun, very like the charged-particle cloud surrounding the Earth which has been recently discovered by Van Allen and his collaborators.

In order to check the hypothesis suggested above, precise measurements of the polarization over a wide spectral range, as well as of the spectral composition of the light, will be important. We are now planning to carry out such measurements during the eclipse of October 2, 1959.

From *Nature* 9 May 1959

## 100 YEARS AGO

Statements have been made in the medical and general Press that the electric waves used in wireless telegraphy are injurious to the operators and produce various diseases, such as conjunctivitis, corneal ulceration, and leukoma. Mr. Marconi writes to the *Times* to deny these suggestions, for which, he says, there is no evidence whatever. He adds:—"During the twelve years or so of our operations we have had to deal with no single case of compensation for any injury of this origin, nor, so far as I can ascertain, has any such injury been suffered. Speaking for myself, I may remark that my own good health has never been better than during the often extended periods when I have been exposed for many hours daily to the conditions now challenged, and in the constant neighbourhood of electrical discharges at our Transatlantic stations, which I believe are the most powerful in the world."

From *Nature* 6 May 1909

50 & 100 YEARS AGO

1% of the dust found in luminous SMGs.

Tamura and colleagues' results<sup>1</sup> are important for a number of reasons. First, if SMGs really are the progenitors of today's most massive elliptical galaxies, which are preferentially found in rich galaxy clusters, the expectation is that they will be found within high-density regions of the high-redshift Universe. This study thus provides powerful support for the 'SMG to giant elliptical' evolutionary pathway, consistent with previous evidence for an excess of SMGs in the high-density regions at high redshift marked out by the most powerful radio galaxies and quasars<sup>8</sup>. Second, it indicates that both the dust-enshrouded and 'transparent' modes of star formation, represented by SMGs and LAEs, respectively, can occur in the same proto-cluster in the same cosmological epoch.

The next obvious step is to obtain spectroscopic

redshifts for the SMGs revealed by the AzTEC image of this field, to confirm how many of them are really physically associated with the  $z = 3.1$  LAE proto-cluster. This will be difficult, as only a subset of SMGs allow sufficient ultraviolet light to escape through the dust for redshifts to be determined with optical telescopes. However, when ALMA becomes operational in the next five years, it will be relatively straightforward to determine the redshifts of known SMGs from their molecular carbon monoxide line emission at millimetre wavelengths. It will also be important to try to invert this study, and search for clusters of LAEs around SMGs that have been revealed by larger-scale, (sub)millimetre surveys<sup>9</sup>. This would allow us to test whether every luminous high-redshift SMG lies within a proto-cluster of lower-mass LAEs. Such multi-waveband studies hold the

key to future progress in our understanding of the formation pathway that led to the galaxies of stars we see around us today. ■

James Dunlop is at the Institute for Astronomy, School of Physics and Astronomy, University of Edinburgh, Royal Observatory, Edinburgh EH9 3HJ, UK.

e-mail: jsd@roe.ac.uk

1. Tamura, Y. *et al.* *Nature* **459**, 61–63 (2009).
2. Koo, D. C. & Kron, R. G. *Publ. Astron. Soc. Pacific* **92**, 537–545 (1980).
3. Iye, M. *et al.* *Nature* **443**, 186–188 (2006).
4. Jimenez, R., Bernardi, M., Haiman, Z., Panter, B. & Heavens, A. F. *Astrophys. J.* **669**, 947–951 (2007).
5. Hughes, D. H. *et al.* *Nature* **394**, 241–247 (1998).
6. Gawiser, E. *et al.* *Astrophys. J.* **671**, 278–284 (2007).
7. Wilson, G. W. *et al.* *Mon. Not. R. Astron. Soc.* **386**, 807–818 (2008).
8. Stevens, J. A. *et al.* *Nature* **425**, 264–267 (2003).
9. Coppin, K. *et al.* *Mon. Not. R. Astron. Soc.* **372**, 1621–1652 (2006).

## TRANSLATION

# Till termination us do part

William Merrick

**Translation of messenger RNA into protein is a complex and intricate process involving several steps and many step-specific protein factors. But one factor — eIF5A — seems to have a hand in every step.**

The process of translation is divided into three main steps: initiation, elongation and termination. Initiation involves the binding of ribosomes — the cell's protein-synthesis machinery — to a specific start site on the messenger RNA sequence. During elongation, the ribosome moves along the mRNA, translating its sequence into a chain of amino acids that are supplied by transfer RNA. Termination occurs when the ribosome and the nascent polypeptide both detach from the mRNA after reaching the end of the protein-coding region on the mRNA. In all three domains of life (bacteria, archaea and eukaryotes), specific protein factors aid each translation step. For example, two universally conserved elongation factors have been identified; their eukaryotic versions are called eEF1A and eEF2. On page 118 of this issue, Saini *et al.*<sup>1</sup> report the existence of a third universal elongation factor, eIF5A, which is unique in that it also acts as an initiation factor and a termination factor.

That eIF5A is an essential protein in yeast is well established. Moreover, much is known about its physical characteristics and biological function<sup>2</sup>. But, despite spectacular advances in the structural analysis of both ribosomes in complex with translation factors and individual factors, the sequential use of eIF5A as a translation factor has not been investigated in detail. Until now, eIF5A has been assumed to be an initiation factor, because when purified it is recovered in association with the initiation factor eIF1A, and it is required for optimal results in a model assay for the initiation step

called methionyl puromycin synthesis<sup>3</sup>.

Several observations set eIF5A apart from other initiation factors. For example, it is the only initiation factor that is more abundant as a free protein than when bound to ribosomes. Also, the lysine residue at position 51 is post-translationally modified to another amino acid, hypusine<sup>4</sup> (N<sup>ε</sup>-(4-amino-2-hydroxybutyl)

lysine), which is not found in any other protein in eukaryotes or archaea<sup>5</sup> and which is essential for eIF5A activity<sup>6</sup>. Curiously, unique post-translational modifications of translation factors have been reported previously for only eEF1A and eEF2 (ref. 2). Finally, sequence alignment analysis<sup>2,7</sup> indicates that eIF5A has evolved from the bacterial elongation factor

## EVOLUTIONARY RELATIONSHIPS BETWEEN BACTERIAL AND EUKARYOTIC TRANSLATION FACTORS

| Bacterial factor      | Function                               | Eukaryotes | Function                               |
|-----------------------|--|------------|--|
| Hypothetical protein* | Unknown                                | eIF1       | Enhances AUG recognition               |
| IF1                   | Enhances AUG recognition               | eIF1A      | Enhances AUG recognition               |
| SelB                  | Binds GTP and selenocysteinyl-tRNA     | eIF2γ      | Binds GTP and met-tRNA <sub>i</sub>    |
| W2                    | A helicase, function not defined       | eIF4A      | DEAD box helicase, subunit of eIF4F    |
| EF-P                  | Elongation                             | eIF5A      | Elongation (and more?)                 |
| IF2                   | Binds GTP and initiator tRNA           | eIF5B      | GTPase, subunit joining                |
| EF1A                  | GTPase, binds aminoacyl-tRNA to A site | eEF1A      | GTPase, binds aminoacyl-tRNA to A site |
| EF2                   | GTPase, ribosome translocation         | eEF2       | GTPase, ribosome translocation         |
| (EF1A)                | GTPase, binds aminoacyl-tRNA to A site | eRF3       | GTPase, termination                    |

All organisms use specific protein factors for each of the three translation steps: initiation (yellow), elongation (blue) and termination (pink). Bacterial and eukaryotic factors with sequence similarity to each other appear together in the same row. They do not, however, necessarily perform the same biological function or facilitate the same reactions in a certain translation step. Exceptions are the three universally conserved elongation factors — including eIF5A, identified by Saini and colleagues<sup>1</sup> — and the initiation factors eIF1 and eIF1A. The colour green indicates factors of unknown function. AUG, initiation codon. (Table adapted from ref. 2.)

\*For example, the hypothetical *E. coli* protein YC1H<sup>7</sup>.



EF-P and not from an initiation factor. Thus, the amino-acid sequence and function<sup>1,2,7,8</sup> of eEF1A, eEF2 and eIF5A are universally maintained. By contrast, the amino-acid sequence and function of many of the polypeptides associated with either the eukaryotic initiation or termination step shows that they have not evolved from equivalent bacterial factors (see table for details).

To directly probe the function of yeast eIF5A, Saini *et al.*<sup>1</sup> used a wide combination of molecular biological and biochemical assays. They find that depletion or inactivation of this factor leads to an increase in both the levels of polysomes — clusters of ribosomes bound to mRNA — and the time it takes ribosomes, after the initiation step, to read the mRNA code and release the nascent polypeptide. What's more, the effects of eIF5A inactivation are similar to those of sordarin, an inhibitor of eEF2.

These observations clearly indicate that eIF5A functions during the elongation step. But why does it also stimulate the initiation<sup>3</sup> and termination<sup>1</sup> steps? One possibility is that eIF5A is required for ribosomes to adopt the most reactive conformation for optimally interacting with factors and aminoacyl-tRNAs that are specifically involved in initiation, elongation and termination.

Two hypothetical models<sup>9,10</sup> — the reciprocating ratchet and the spring and ratchet — provide an account of how the mRNA and growing polypeptide chain are moved through the ribosome. According to these models, throughout the elongation step, the large and small ribosomal subunits move back and forth relative to one another, facilitating the movement of the mRNA and the growing polypeptide (peptidyl-tRNA) along the three consecutive sites on the ribosome — A (tRNA-binding), P (peptide-bond formation) and E (exit). Thus, faithful translation of the mRNA sequence into a polypeptide chain is ensured. As eIF5A also stimulates the initiation and termination steps, it is likely that this factor is bound to the ribosome while an initiator tRNA or the peptidyl-tRNA is bound to the P site, making the tRNA more reactive with an incoming aminoacyl-tRNA or termination factor at the A site.

Given that, at a relative molecular mass ( $M_r$ ) of 15,000, eIF5A is much smaller than a ribosome — which has an  $M_r$  of roughly 4 million — it is unlikely that its effect on ribosome conformation can be detected experimentally. Nonetheless, data obtained through cryo-electron microscopy and three-dimensional reconstructions show<sup>11</sup> that, when bound to the 40S subunit of ribosomes, the initiation factors eIF1 and eIF1A (both about the same size as eIF5A) dramatically change the shape of this subunit, although the factors themselves cannot be visualized. So it might be possible to similarly investigate whether eIF5A alters the shape or subunit orientation of eukaryotic ribosomes, without necessarily visualizing it.

Saini and colleagues' results do not just establish a role for eIF5A as a major player

during the elongation step. In eIF5A, they also add a useful tool to the experimental tool box, which should allow high-resolution probing of the alternating states that eukaryotic ribosomes might assume in order to accurately and efficiently catalyse the various steps of translation.

William Merrick is in the Department of Biochemistry, Case Western Reserve University School of Medicine, Cleveland, Ohio 44106, USA. e-mail: wcm2@case.edu

1. Saini, P., Eyler, D. E., Green, R. & Dever, T. E. *Nature* **459**, 118–121 (2009).

2. Hershey, J. W. B. & Merrick, W. C. In *Translational Control of Gene Expression* (eds Sonenberg, N., Hershey, J. W. B. & Mathews, M. B.) 33–88 (Cold Spring Harbor Press, 2000).
3. Kemper, W. M., Berry, K. W. & Merrick, W. C. *J. Biol. Chem.* **251**, 5551–5557 (1976).
4. Cooper, H. L. *et al. Proc. Natl Acad. Sci. USA* **80**, 1854–1857 (1983).
5. Gordon, E. D. *et al. J. Biol. Chem.* **262**, 16585–16589 (1987).
6. Park, M. H. *et al. J. Biol. Chem.* **266**, 7988–7994 (1991).
7. Kyrpides, N. C. & Woese, C. R. *Proc. Natl Acad. Sci. USA* **95**, 224–228 (1998).
8. Labib, K., Tercero, J. A. & Diffley, J. F. *Science* **288**, 1643–1647 (2000).
9. Woese, C. *Nature* **226**, 817–820 (1970).
10. Moran, S. J., Flanagan, J. F. 4th, Namy, O., Stuart, D. I., Brierly, I. & Gilbert, R. J. *Structure* **16**, 664–672 (2008).
11. Passmore, L. A. *et al. Mol. Cell* **26**, 41–50 (2007).

## MECHANOCHEMISTRY

# Polymers react to stress

Christoph Weder

**The latest polymers are chameleon-like: they change colour on deformation. The transduction mechanism underpinning this effect could be used to make polymers that respond in many other ways to mechanical stress.**

Imagine a polymer that could send a warning signal if stressed close to the point of mechanical failure. Or one that actually becomes stronger under load. Or even one that heals itself after being damaged. On page 68 of this issue, Davis *et al.*<sup>1</sup> report a breakthrough that could enable all three possibilities: polymers in which the application of an external force activates preprogrammed chemical reactions, which in turn cause desired responses. The authors used this approach to make mechanochromic polymers — materials that change colour on deformation.

There is currently great interest in polymers whose properties change in response to stimuli such as chemicals, heat, light or electricity, because of their potential use in applications ranging from camouflage systems to artificial muscles to drug delivery<sup>2</sup>. Many such materials have been made, but comparably few polymers have been developed that respond in a useful way to mechanical stress<sup>3</sup>. These include materials that change their absorption or fluorescence colour on deformation, so providing visible warning signs before mechanical failure occurs<sup>4–6</sup>. The colour changes are caused by several mechanisms, including variations in the molecular interactions between dye molecules integrated into the polymers<sup>4</sup>, or in the molecular conformations of integrated dyes<sup>5</sup>. Alternatively, in photonic-bandgap materials (through which only light of a specific wavelength can propagate), deformation of the sample changes the distance between particles in the material's lattice structure, which in turn changes the wavelength of light that passes through. But these mechanisms are fundamentally different from those of mechanically responsive materials found in nature, essen-

tially all of which translate macroscopic forces into chemical reactions<sup>7</sup>.

Davis *et al.*<sup>1</sup> set out to emulate nature's approach by making artificial polymers in which mechanical stress provides the activation energy for specific chemical reactions. The mechanochemically reactive unit — the mechanophore — used by the authors is a 'spiropyran' dye (see Fig. 1a on page 69). Spiroyrans change colour when exposed to light or heat, sometimes converting from colourless to highly coloured states, and they are used in applications such as self-darkening spectacles. These optical changes occur because light or heat activates a reaction in which a weak bond in the spiropyran is broken, causing the molecules to rearrange into a coloured product. The colour changes can also be triggered by grinding the compounds<sup>8</sup>, demonstrating that such compounds have mechanochromic characteristics.

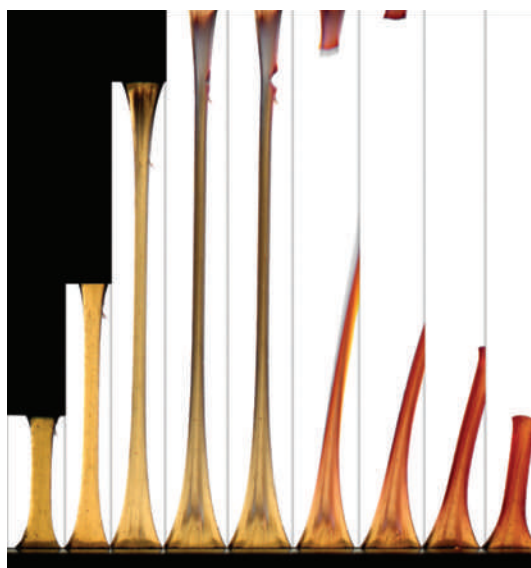
Seeking to make polymers that have mechanochromic properties, Moore and Sottos (the lead authors of the paper by Davis *et al.*<sup>1</sup>) have previously incorporated several mechanophores — including a spiropyran — into polymers<sup>9,10</sup>. Assuming that efficient force transfer between the macromolecules and the mechanophore would be essential, they developed several approaches for covalently connecting spiroyrans to polymers. One route involved attaching two chemical groups to a spiropyran core; the groups acted as initiation points from which polymer chains were grown. In this way, the authors prepared a polymer in which each molecule contains a spiropyran at its centre. They then showed that solutions of the polymer change from being colourless to pink on exposure to ultrasound<sup>10</sup>. This

happens because ultrasound-generated shear forces are funnelled through the polymer chains to the spiropyran, which responds by reacting as described previously. Other research groups have also shown that mechanoreceptors in polymers can break if the mechanophores contain sufficiently weak bonds. This was demonstrated recently by the reported ultrasound-induced activation of certain catalysts<sup>11</sup>, for example.

So far, so good. But in most real-world situations, forces are not applied to polymer solutions by ultrasound; rather, they are applied to solids by stretching or compression. Furthermore, when solid polymers are exposed to excessive stress, the normal mechanochemical response is for the polymer molecules to break at random points along the chain<sup>12</sup>. A key question therefore remained: if solid spiropyran-containing polymers are placed under stress, does that stress break the mechanophores? Davis *et al.*<sup>1</sup> now show that it does. When they subjected spiropyran-containing polymers to stress, either by stretching a rubbery poly(methyl acrylate) polymer or by compressing a glassy poly(methyl methacrylate) polymer, the colour and fluorescence of the materials changed, indicating that the spiropyran had reacted (Fig. 1).

Using a combination of experimental and theoretical models, Davis *et al.* showed that the underlying mechanism of the colour change is undeniably stress-induced reaction of the spiropyran upon irreversible deformation of the material. They further showed that the connectivity between the mechanophore and the polymer is important: placing two polymer chains on opposite sides of the spiropyran allows the maximum transfer of force from the polymer chains, whereas attaching chains to the same side of the spiropyran, or using only one polymer chain, prevents or limits transfer to the desired breaking point. The authors thus report the first set of guiding principles for the design of solid mechanochemical polymers. Nevertheless, further studies are required to work out just how selective the mechanochemical transduction really is. Although the authors' computational studies suggest that chain scission occurs only at the spiropyran, the possibility that rupture also occurs in other parts of the polymer chains cannot be ruled out.

Davis and colleagues' approach to making mechanochemical polymers should be readily adaptable to other polymer systems. But the development of materials that use spiropyran mechanophores as built-in strain sensors will require several practical problems to be solved. For example, the colour change in the polymers described<sup>1</sup> doesn't occur only in response to deformation; it also does so in response to heat and light. Furthermore, the ring-opening reaction of spiropyran is reversed by exposure to light. The resulting loss of colour



**Figure 1 | Red for danger.** Davis *et al.*<sup>1</sup> have prepared mechanochemical polymers that change colour under mechanical stress. When a sample of one of these polymers is stretched, the material turns red at the point where irreversible deformation (tearing) starts to occur, as shown in this sequence of pictures. Once the material snaps, the red colour is apparent throughout the sample. The sample on the left is 7.4 mm long.

would clearly be a problem if the polymers are to be used as damage sensors.

The importance of Davis and colleagues' work goes far beyond the reported mechanochemical effect — it may become a milestone in polymer science, because the authors' general design concept can probably be adapted to make a plethora of materials that translate mechanical stress into all kinds of useful chemical reactions. The authors propose<sup>1</sup> that polymerization and molecular-crosslinking reactions could be triggered upon deformation. This could be achieved by mechanochemical processes that release reaction initiators or

catalysts, and might be used to create self-healing materials. The reverse effect — cutting across molecules at predetermined breaking points — could be used to create polymers that have stress-activated fuses.

Other applications include materials that have stress-controlled drug-release properties, which could be used in tissues that are naturally subjected to mechanical stress; and substrates for cell culture that contain dimples in which cells won't grow, because the properties of those regions were mechanochemically altered when the dimples were stamped into the substrate. Scientists from many disciplines will be restricted only by their imaginations when it comes to finding ways of using mechanically responsive polymers for their specific needs. And studies of well-defined artificial systems might contribute to our often rather limited understanding of mechanochemical transduction in biological systems. ■

Christoph Weder is at the Adolphe Merkle Institute, University of Fribourg, CH-1700 Fribourg, Switzerland, and in the Department of Macromolecular Science and Engineering, Case Western Reserve University, Cleveland 44106-7202, USA.  
e-mail: christoph.weder@unifr.ch

1. Davis, D. A. *et al.* *Nature* **459**, 68–72 (2009).
2. Shahinpoor, M. & Schneider, H.-J. (eds) *Intelligent Materials* (Roy. Soc. Chem. Press, 2008).
3. Comrie, J. E. & Huck, W. T. S. *Macromol. Rapid Commun.* **29**, 539–546 (2008).
4. Löwe, C. & Weder, C. *Adv. Mater.* **14**, 1625–1629 (2002).
5. Nallicheri, R. A. & Rubner, M. F. *Macromolecules* **24**, 517–525 (1991).
6. Foulger, S. H. *et al.* *Adv. Mater.* **13**, 1898–1901 (2001).
7. Heinrichs, A. *Nature Rev. Mol. Cell Biol.* **10**, 163 (2009).
8. Tipikin, D. S. *Russ. J. Phys. Chem.* **75**, 1720–1722 (2001).
9. Hickenboth, C. R. *et al.* *Nature* **446**, 423–427 (2007).
10. Potisek, S. L. *et al.* *J. Am. Chem. Soc.* **129**, 13808–13809 (2007).
11. Piermattei, A. *et al.* *Nature Chem.* **1**, 103–137 (2009).
12. Beyer, M. K. & Clausen-Schaumann, H. *Chem. Rev.* **105**, 2921–2948 (2005).

## CELL BIOLOGY

# Arrest by ribosome

Sébastien Ferreira-Cerca and Ed Hurt

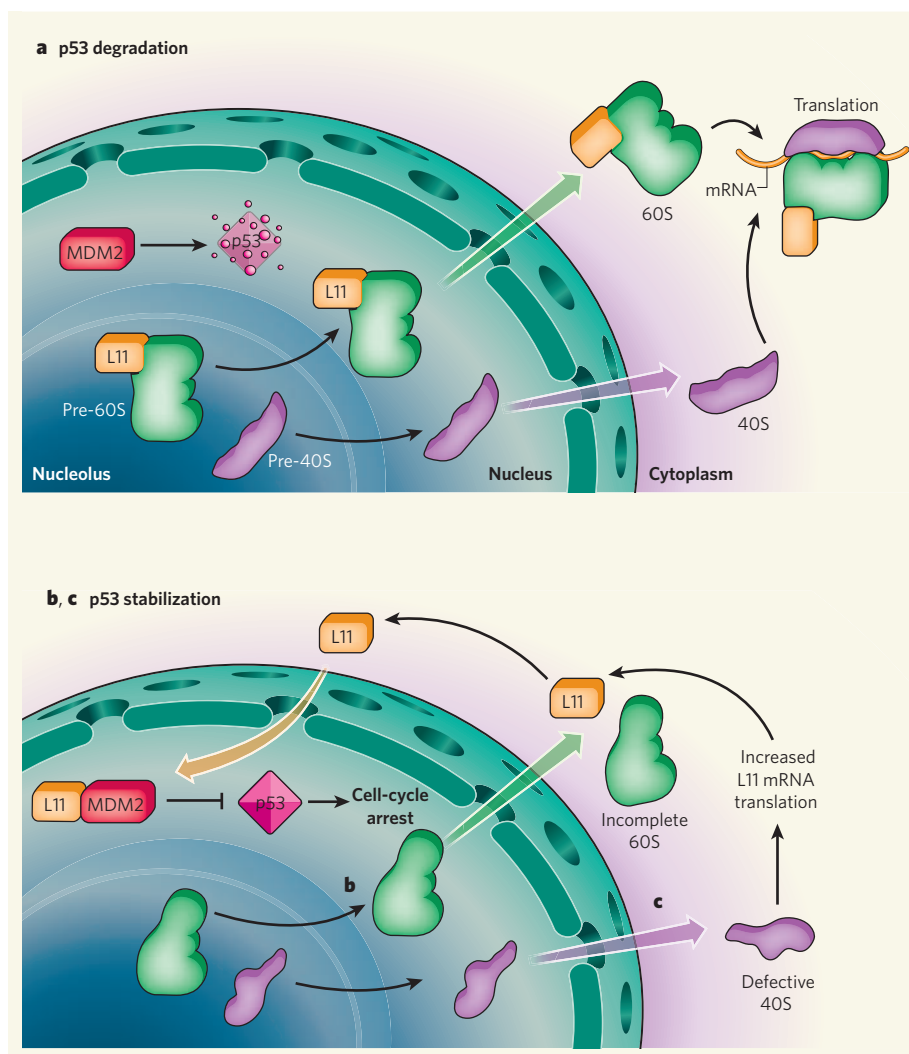
**Impaired assembly of cells' protein-synthesis factories, the ribosomes, can cause cell-cycle arrest and disease. This finding emphasizes the close link between cell proliferation and ribosome formation.**

Protein synthesis is mediated by complexes of RNA and protein known as ribosomes. Ribosome biogenesis is complicated, involving some 150 non-ribosomal factors and 100 small non-coding RNAs<sup>1,2</sup>. It is also the most energy-consuming process in growing cells, and so requires extensive regulation and coordination. Pressing questions are how ribosome synthesis is regulated, how it links to cell proliferation, and how it responds to environmental

cues such as nutrient availability and stress. Writing in *Nature Cell Biology*, Fumagalli *et al.*<sup>3</sup> provide insight not only into the molecular mechanisms that connect ribosome biogenesis and cell proliferation, but also into underlying human diseases associated with defective ribosome synthesis — ribosomopathies<sup>4,5</sup>.

Eukaryotic organisms (such as yeast, plants and animals) have two ribosomal subunits, 60S and 40S, which are assembled in the





**Figure 1 | Ribosomal stress and p53 stabilization.** **a**, In normal cells, after initial biogenesis of the 60S and 40S ribosomal subunits in the nucleolus, they are exported out of the nucleus, where the mature subunits come together to mediate mRNA translation. Under these conditions, MDM2 is free to mediate p53 degradation in the nucleus. **b**, **c**, In mutant cells with impaired ribosome assembly, the L11 component of 60S can cause cell-cycle arrest by binding to MDM2 and preventing it from degrading p53. Free L11 can be generated in two ways: through defective 60S biogenesis (**b**); or, as Fumagalli *et al.*<sup>3</sup> show, by increased translation of its mRNA, when 40S biogenesis is defective (**c**).

nucleolus — a nuclear structure that specializes in ribosome production — before being exported to the cytoplasm. Previous work has shown that, on disruption of the nucleolus, ribosome synthesis is impaired, leading to a halt in the cell cycle<sup>6–8</sup>. Unexpectedly, such nucleolar stress induces and stabilizes p53, a key regulator of cell proliferation. In fact, on disintegration of the nucleolus, several ribosomal proteins are released into the nucleoplasm and bind to the enzyme MDM2. This E3 ubiquitin ligase normally promotes p53 degradation, but, when bound to ribosomal proteins, it no longer functions properly. As a result, p53 levels rise, causing cell-cycle arrest, apoptotic cell death and/or senescence<sup>6–8</sup>.

Fumagalli *et al.*<sup>3</sup> specifically disrupt the formation of the 40S subunit using RNA interference, without disrupting the nucleolus. In

agreement with another recent study<sup>9</sup>, they find that, even in the presence of an intact nucleolus, both p53 stabilization and the subsequent cell-cycle arrest occur. Strikingly, regardless of whether biogenesis of the 60S or the 40S subunit is impaired, it is a 60S protein component, L11, that mediates p53 stabilization by binding to MDM2. But, whereas defects in 60S biogenesis are expected to lead to an increase in free L11, how can impaired formation of 40S yield this protein?

The authors<sup>3</sup> show that disruption of 40S biogenesis activates the translation of a group of messenger RNAs called 5'-TOP mRNAs — including the L11 mRNA — which contain a sequence motif of several pyrimidine bases in their 5'-untranslated regions. So free L11 can be generated in two ways: its mRNA translation increases when 40S biogenesis

is impaired; and it is independently released into the nucleoplasm when 60S biogenesis is defective (Fig. 1). It remains to be seen how the translation machinery senses disruption in 40S biogenesis.

Fumagalli and colleagues' observations<sup>3</sup> unravel an exciting direct link between ribosomal stress and p53 induction, which will undoubtedly stimulate further investigation. Earlier work<sup>6,7</sup> has indicated that, as well as L11, other ribosomal proteins from both 40S and 60S can bind to MDM2 to stabilize p53. With L11 having the pivotal role in p53 stabilization, what contribution do these other proteins make? Moreover, because L11 assembly onto the nascent 60S subunit occurs early on in the nucleolus<sup>1,2</sup>, how would defects in the biogenesis of this subunit later on, downstream of L11 assembly, generate the p53-stabilizing form of L11?

And can Fumagalli and co-workers' findings help in understanding the ribosomopathies at the molecular level? Notably, genes encoding protein components of 60S, including L5 and L11 (ref. 10), are mutated in some patients with Diamond-Blackfan anaemia. It is important to establish whether p53 is induced in the cells of such patients. If so, how could p53 be stabilized when L11 is mutated? If p53 is indeed stabilized in these patients, the combined effects of altered p53 levels and imbalance in ribosome assembly may cause this ribosomopathic anaemia. Systematically analysing the factors involved in the crosstalk between p53 stabilization and ribosome stress will undoubtedly aid in identifying specific molecular culprits.

If p53 stabilization does turn out to be a hallmark of ribosomopathies, targeted drug therapies in patients with these conditions will require keeping a balance between sufficient ribosome formation and adequate p53 safeguarding functions. With our growing understanding of ribosome synthesis and function, and characterization of the molecular basis of ribosomopathies, the development of targeted treatments for these diseases is looking closer than ever.

Sébastien Ferreira-Cerca and Ed Hurt are at the University of Heidelberg, Biochemistry Center (BZH), Im Neuenheimer Feld 328, D-69120 Heidelberg, Germany.  
e-mail: ed.hurt@bzh.uni-heidelberg.de

- Henras, A. K. *et al.* *Cell. Mol. Life Sci.* **65**, 2334–2359 (2008).
- Tschochner, H. & Hurt, E. *Trends Cell. Biol.* **13**, 255–263 (2003).
- Fumagalli, S. *et al.* *Nature Cell Biol.* **11**, 501–508 (2009).
- Ellis, S. R. & Lipton, J. M. *Curr. Top. Dev. Biol.* **82**, 217–241 (2008).
- Luzzatto, L. & Karadimitris, A. *Nature Genet.* **19**, 6–7 (1998).
- Chen, D. *et al.* *Oncogene* **26**, 5029–5037 (2007).
- Lindstrom, M. S., Deisenroth, C. & Zhang, Y. *Cell Cycle* **6**, 434–437 (2007).
- Mayer, C. & Grummt, I. *Cell Cycle* **4**, 1036–1038 (2005).
- Barkic, M. *et al.* *Mol. Cell. Biol.* **29**, 2489–2504 (2009).
- Gazda, H. T. *et al.* *Am. J. Hum. Genet.* **83**, 769–780 (2008).

# The formation of the first stars and galaxies

Volker Bromm<sup>1</sup>, Naoki Yoshida<sup>2</sup>, Lars Hernquist<sup>3</sup> & Christopher F. McKee<sup>4</sup>

**Observations made using large ground-based and space-borne telescopes have probed cosmic history from the present day to a time when the Universe was less than one-tenth of its present age. Earlier still lies the remaining frontier, where the first stars, galaxies and massive black holes formed. They fundamentally transformed the early Universe by endowing it with the first sources of light and chemical elements beyond the primordial hydrogen and helium produced in the Big Bang. The interplay of theory and upcoming observations promises to answer the key open questions in this emerging field.**

The formation of the first stars and galaxies at the end of the cosmic ‘dark ages’ is one of the central problems in modern cosmology<sup>1–3</sup>. It is thought that during this epoch the Universe was transformed from its simple initial state into a complex, hierarchical system, through the growth of structure in the dark matter, by the input of heavy elements from the first stars, and by energy injection from these stars and from the first black holes<sup>4,5</sup>. An important milestone in our understanding was reached after the introduction of the now standard cold dark-matter (CDM) model of cosmic evolution, which posits that structure grew hierarchically, such that small objects formed first and then merged to form increasingly larger systems<sup>6</sup>. Within this model, dark-matter ‘minihaloes’ (see below), forming a few hundred million years after the Big Bang, were identified as the sites where the first stars formed<sup>7</sup>. Building on this general framework, and relying on the development of efficient new computational tools, the fragmentation properties of primordial gas inside such minihaloes were investigated with numerical simulations, leading to the result that the first stars, so-called population III, were predominantly very massive<sup>4,8</sup> (see Box 1 for the terminology used here). Recently, the frontier has progressed to the next step in the hierarchical build-up of structure, to the emergence of the first galaxies whose formation took place after the first stars had formed and affected their common environment. It is very timely to review our current understanding and remaining challenges, as we are just entering an exciting period of discovery, in which new observational probes are becoming available and advances in super-computer technology enable ever more realistic theoretical predictions.

We begin with the formation of the first stars, discussing the physics underlying the prediction that they were very massive, and how this picture would be modified if the dark matter exhibited non-standard properties on small scales. We next address the feedback effects from the first stars, with one main result being that such feedback might delay subsequent star formation by up to  $\sim 10^8$  years. Proceeding to the assembly of the first galaxies, we discuss the important role of turbulence and supernova feedback during their formation. Intriguingly, the cold accretion streams that feed the turbulence in the centres of the primordial galaxies are reminiscent of the recently proposed new model for galaxy formation, in which such cold streams are invoked to explain the build-up of massive galaxies at more recent cosmic times in a smooth, rather than merger-driven, fashion<sup>9</sup>. We conclude with an outlook into the likely key developments over the next decade.

## Formation of the first stars

Whereas dark-matter haloes can originate through the action of gravity alone, the formation of luminous objects, such as stars and

galaxies, is a much more complicated process. For star formation to begin, a sufficient amount of cold dense gas must accumulate in a dark halo. In the early Universe, the primordial gas could not efficiently cool radiatively because atoms have excitation energies that are too high, and molecules, which have accessible rotational energies, are very rare. Trace amounts of molecular hydrogen ( $H_2$ ) can be produced via a sequence of reactions,  $H + e^- \rightarrow H^- + \gamma$  (where  $\gamma$  indicates a photon), followed by  $H^- + H \rightarrow H_2 + e^-$  (where  $e^-$  indicates an electron), and under the proper conditions this allows the gas to cool and eventually condense to form stars<sup>10</sup>.

Numerical simulations<sup>11–13</sup> starting from cosmological initial conditions show that primordial gas clouds formed in dark-matter haloes with virial temperature  $\sim 1,000$  K and mass  $\sim 10^6 M_\odot$ .

### Box 1 | Definitions and terminology

We establish here a convention for terminology used in this review. Population III stars are those that initially contain no elements heavier than helium (‘metals’ in the parlance of astronomers) other than the lithium produced in the Big Bang. Such stars can be divided into first generation stars (population III.1), which form from initial conditions determined entirely by cosmological parameters, and second generation stars (population III.2), which originate from material that was influenced by earlier star formation<sup>25</sup>. According to theory, population III.1 stars formed when almost completely neutral primordial gas collapsed into dark-matter minihaloes, whereas one important class of population III.2 stars formed from gas that was photoionized before the onset of gravitational runaway collapse<sup>33</sup>. Simply put, population III.1 stars are locally the very first luminous objects, whereas population III.2 stars are those metal-free stars formed from gas that was already affected by previous generations of stars. Population II stars have enough metals to affect their formation and/or their evolution. Such stars are classified<sup>100</sup> according to their iron/hydrogen ratio as extremely metal poor for metallicities  $10^{-4} < Z/Z_\odot < 10^{-3}$ , ultra-metal poor for  $10^{-5} < Z/Z_\odot < 10^{-4}$ , and hyper-metal poor for  $10^{-6} < Z/Z_\odot < 10^{-5}$ . Because we know so little about the first galaxies, it is difficult to establish a precise terminology for them. A galaxy is a system of many stars and gas that is gravitationally bound in a dark-matter halo. We define a ‘first galaxy’ as one composed of the very first system of stars to be gravitationally bound in a dark-matter halo. Such stars could be population III or population II stars with very low metallicities—extremely metal poor or below, according to recent numerical simulations<sup>63,64</sup>. The gas in such galaxies should have similarly low metallicities. Current theory predicts that population III.1 stars are formed in isolation in minihaloes and therefore will not be in galaxies.

<sup>1</sup>Astronomy Department, University of Texas, 2511 Speedway, Austin, Texas 78712, USA. <sup>2</sup>Institute for the Physics and Mathematics of the Universe, University of Tokyo, Kashiwa, Chiba 277-8568, Japan. <sup>3</sup>Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA. <sup>4</sup>Departments of Physics and Astronomy, University of California, Berkeley, California 94720, USA.



(so-called ‘minihaloes’;  $M_{\odot}$ , solar mass). In the standard CDM model, the minihaloes that were the first sites for star formation are expected to be in place at redshift  $z \approx 20$ –30, when the age of the Universe was just a few hundred million years<sup>14</sup>. These systems correspond to  $(3\text{--}4)\sigma$  peaks in the cosmic density field, which is statistically described as a Gaussian random field. Such high-density peaks are expected to be strongly clustered<sup>15</sup>, and thus feedback effects from the first stars are important in determining the fate of the surrounding primordial gas clouds. It is very likely that only one star can be formed within a gas cloud, because the far-ultraviolet radiation from a single massive star is sufficient to destroy all the  $\text{H}_2$  in the parent gas cloud<sup>16,17</sup>. In principle, a cloud that formed one of the first stars could fragment into a binary or multiple star system<sup>18,19</sup>, but simulations based on self-consistent cosmological initial conditions do not show this<sup>20</sup>. Although the exact number of stars per cloud cannot be easily determined, the number is expected to be small, so that minihaloes will not be galaxies (see Box 1).

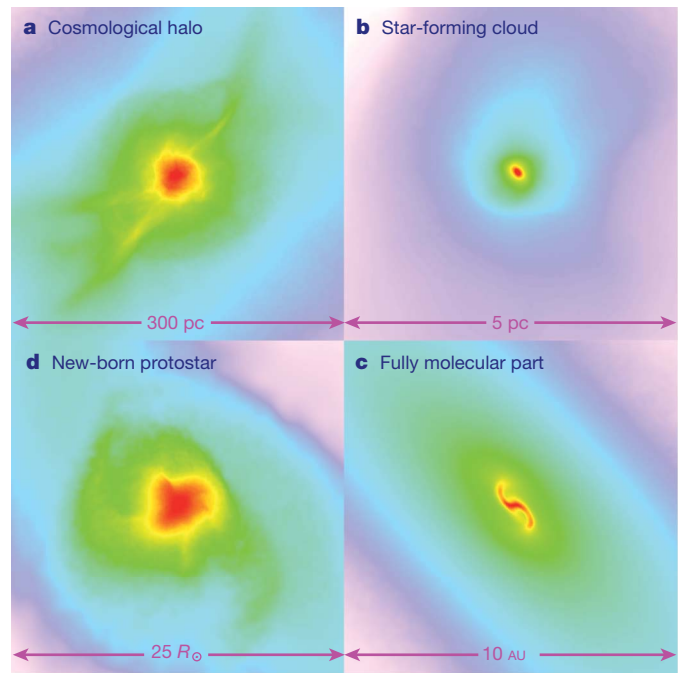
Primordial gas clouds undergo runaway collapse when sufficient mass is accumulated at the centre of a minihalo. The minimum mass at the onset of collapse is determined by the Jeans mass (more precisely, the Bonnor–Ebert mass), which can be written as:

$$M_J \approx 500 M_{\odot} \left( \frac{T}{200} \right)^{3/2} \left( \frac{n}{10^4} \right)^{-1/2} \quad (1)$$

for an atomic gas with temperature  $T$  (in K) and particle number density  $n$  (in  $\text{cm}^{-3}$ ). The characteristic temperature is set by the energy separation of the lowest-lying rotational levels of the trace amounts of  $\text{H}_2$ , and the characteristic density corresponds to the thermalization of these levels, above which cooling becomes less efficient<sup>12</sup>. A number of atomic and molecular processes are involved in the subsequent evolution of a gravitationally collapsing gas. It has been suggested that a complex interplay between chemistry, radiative cooling and hydrodynamics leads to fragmentation of the cloud<sup>21</sup>, but vigorous fragmentation is not observed even in extremely high-resolution cosmological simulations<sup>11–13,20,22</sup>. Interestingly, however, simulations starting from non-cosmological initial conditions have yielded multiple cloud cores<sup>19,23</sup>. It appears that a high initial degree of spin in the gas eventually leads to the formation of a disk and its subsequent break-up. It remains to be seen whether such conditions occur from realistic cosmological initial conditions.

Although the mass triggering the first runaway collapse is well-determined, it provides only a rough estimate of the mass of the star(s) to be formed. Standard star-formation theory predicts that a tiny protostar forms first and subsequently grows by accreting the surrounding gas to become a massive star. Indeed, the highest-resolution simulations of first-star formation verify that this also occurs cosmologically<sup>20</sup> (Fig. 1). However, the ultimate mass of the star is determined both by the mass of the cloud out of which it forms and by a number of feedback processes that occur during the evolution of the protostar. In numerical simulations, the final mass of a population III star is usually estimated from the density distribution and velocity field of the surrounding gas when the first protostellar fragment forms, but this may well be inaccurate even in the absence of protostellar feedback. Whereas protostellar feedback effects are well studied in the context of the formation of contemporary stars<sup>24</sup>, they differ in several important respects in primordial stars<sup>25</sup>.

First, primordial gas does not contain dust grains. As a result, radiative forces on the gas are much weaker. Second, it is generally assumed that magnetic fields are not important in primordial gas because, unless exotic mechanisms are invoked, the amplitudes of magnetic fields generated in the early Universe are so small that they never become dynamically significant in primordial star-forming gas<sup>26</sup>. Magnetic fields have at least two important effects in contemporary star formation: they reduce the angular momentum of the gas out of which stars form, and they drive powerful outflows that disperse a significant fraction of the parent cloud. It is likely that the pre-stellar gas has more angular momentum in the primordial case, and this is borne out by cosmological simulations. Third, primordial stars are

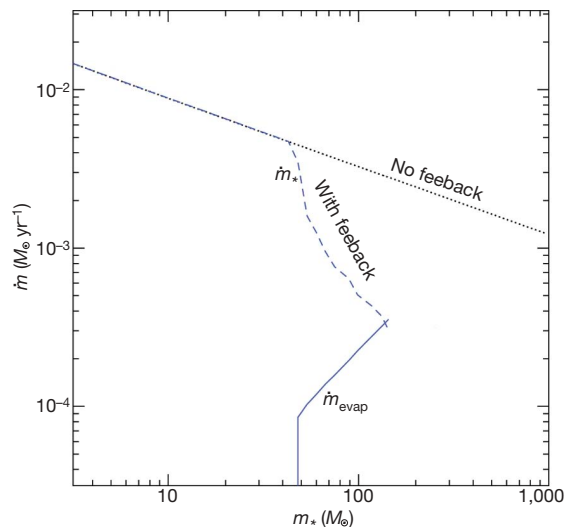


**Figure 1 | Projected gas distribution around a primordial protostar.** Shown is the gas density (colour-coded so that red denotes highest density) of a single object on different spatial scales. **a**, The large-scale gas distribution around the cosmological minihalo; **b**, a self-gravitating, star-forming cloud; **c**, the central part of the fully molecular core; and **d**, the final protostar. Reproduced by permission of the AAAS (from ref. 20).

much hotter than contemporary stars of the same mass, resulting in significantly greater ionizing luminosities<sup>27</sup>.

State-of-the-art numerical simulations of the formation of the first (population III.1) stars represent a computational *tour de force*, in which the collapse is followed from cosmological (comoving megaparsec) scales down to protostellar (sub-astronomical-unit) scales, revealing the entire formation process of a protostar. However, further growth of the protostar cannot be followed accurately without implementing additional radiative physics. For now, inferring the subsequent evolution of the protostar requires approximate analytic calculations. By generalizing a theory for contemporary massive-star formation<sup>28</sup>, it is possible to approximately reproduce the initial conditions found in the simulations and to then predict the growth of the accretion disk around the star<sup>29</sup>. Several feedback effects determine the final mass of a first star<sup>25</sup>: photodissociation of  $\text{H}_2$  in the accreting gas reduces the cooling rate, but does not stop accretion. Lyman- $\alpha$  radiation pressure can reverse the infall in the polar regions when the protostar grows to  $20\text{--}30 M_{\odot}$ , but cannot significantly reduce the accretion rate. The expansion of the  $\text{H II}$  region produced by the large flux of ionizing radiation can significantly reduce the accretion rate when the protostar reaches  $50\text{--}100 M_{\odot}$ , but accretion can continue in the equatorial plane. Finally, photoevaporation-driven mass loss from the disk<sup>30</sup> stops the accretion and fixes the mass of the star (see Fig. 2). The final mass depends on the entropy and angular momentum of the pre-stellar gas; for reasonable conditions, the mass spans  $60\text{--}300 M_{\odot}$ .

A variety of physical processes can affect and possibly substantially alter the picture outlined above. Magnetic fields generated through the magneto-rotational instability may become important in the protostellar disk<sup>31</sup>, although their strength is uncertain, and may play an important role in the accretion phase<sup>18</sup>. Cosmic rays and other external ionization sources, if they existed in the early Universe, could significantly affect the evolution of primordial gas<sup>32</sup>. A partially ionized gas cools more efficiently because the abundant electrons promote  $\text{H}_2$  formation. Such a gas cools to slightly lower temperatures than a neutral gas can, accentuating the fractionation of D into HD so that cooling by HD molecules becomes important<sup>33–36</sup>.



**Figure 2 | Feedback-limited accretion.** Change in mass ( $\dot{m}$ ) versus protostellar mass ( $m_*$ ) for a number of key processes. The protostellar accretion rate ( $\dot{m}_*$ ) is shown in the cases of ‘no feedback’ (black dotted line) and ‘with feedback’ (blue dashed line). Even as an H II region is built up, accretion continues through an accretion disk, which is eventually destroyed via photoevaporation. Also shown is the corresponding rate ( $\dot{m}_{\text{evap}}$ ; blue solid line). The intersection of the blue dashed and solid curves determines the final population III mass. Reproduced by permission of the AAS (from ref. 25).

More significant modifications to the standard model result if the properties of the dark matter are different from those assumed above (see Fig. 3). A key assumption in the standard model is that the dark matter interacts with the baryons only via gravity. However, dark matter can indirectly affect the dynamics of a pre-stellar gas. A popular candidate for CDM is the neutralino, for which the self-annihilation cross-section is large. Neutralino dark matter is thus expected to pair-annihilate in very dense regions, producing high-energy particles such as pions and electron–positron pairs and high-energy photons. These annihilation products may effectively heat collapsing primordial gas clouds when the density is sufficiently high, thereby arresting the collapse<sup>37</sup>. Calculation of the structure of stars with dark-matter annihilation suggest that they can undergo a phase of evolution in which they have temperatures of 4,000–10,000 K, well below those for conventional population III stars<sup>38,39</sup>. The magnitude of this effect depends sensitively on details such as the dark-matter concentration and the final products of neutralino annihilation. Furthermore, calculations to date have assumed spherical symmetry, whereas it is possible that the angular momentum of both the baryons (which leads to the formation of an accretion disk<sup>29</sup>) and of the dark matter could

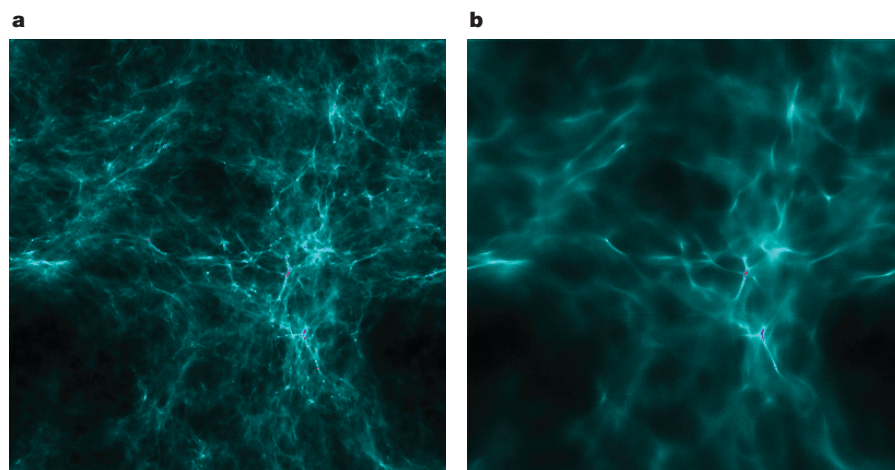
significantly impede the build-up of the high dark-matter densities required to power the stellar luminosity via dark-matter annihilation. Nevertheless, if neutralinos are detected in the appropriate mass range<sup>40</sup>, early star formation models may need to include the effect of dark-matter annihilation.

### Feedback from the first stars

Some of the feedback processes described above that affect the formation of individual stars also influence primordial star formation on large scales. The enormous fluxes of ionizing radiation and H<sub>2</sub>-dissociating Lyman–Werner radiation emitted by massive population III stars<sup>27,41</sup> dramatically influence their surroundings, heating and ionizing the gas within a few kiloparsecs of the progenitor and destroying the H<sub>2</sub> within a somewhat larger region<sup>17,33,42–44</sup>. Moreover, the Lyman–Werner radiation emitted by the first stars could propagate across cosmological distances, allowing the build-up of a pervasive Lyman–Werner background radiation field<sup>45,46</sup>. The effect of radiation from the first stars on their local surroundings has important implications for the numbers and types of population III stars that form. The photoheating of gas in the minihaloes hosting population III.1 stars drives strong outflows, lowering the density of the gas in the minihaloes and delaying subsequent star formation by up to 100 Myr (ref. 47). Furthermore, neighbouring minihaloes may be photoevaporated, delaying star formation in such systems as well<sup>48–50</sup>. The photodissociation of molecules by Lyman–Werner photons emitted from local star-forming regions will, in general, act to delay star formation by destroying the main coolants that allow the gas to collapse and form stars<sup>51</sup>.

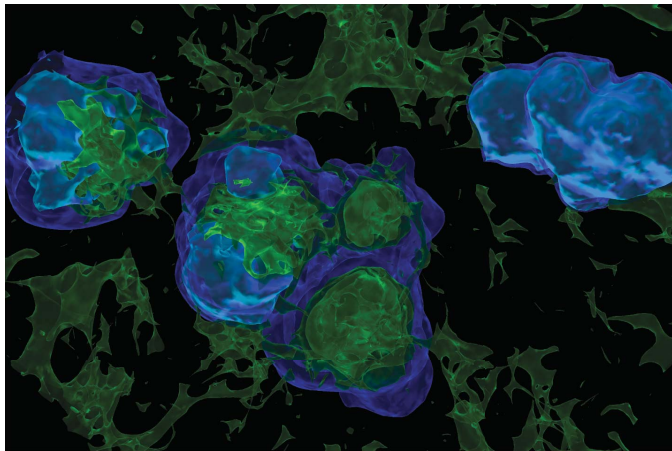
The photoionization of primordial gas can, however, also stimulate star formation by fostering the production of abundant molecules within the relic H II regions surrounding the remnants of population III.1 stars<sup>44,47,52,53</sup> (see Fig. 4). It is still debated whether this radiative feedback is positive or negative in terms of its overall impact on the cosmic star formation rate<sup>54</sup>. However, some robust conclusions have emerged from the recent simulations. First, the Lyman–Werner feedback is much less ‘suicidal’ than was originally thought<sup>55</sup>. It is now believed that star formation in neighbouring minihaloes is not completely suppressed, but merely delayed. Second, the ionizing radiation from the first stars is initially very disruptive because it substantially decreases the density in the host minihalo. This effect leads to the substantial gap between the formation of the first and second generations of stars. In each region of space, the drama of ‘first light’ thus occurred in two clearly separated stages.

Most of the work on the evolution of population III stars and on the supernovae they produce has been based on the assumption that the stars are not rotating<sup>56</sup>. For initial stellar masses in the range  $25M_{\odot} \lesssim M_* \lesssim 140M_{\odot}$  and  $M_* \gtrsim 260M_{\odot}$ , population III stars end their lives by collapsing into black holes with relatively little ejection of heavy



**Figure 3 | Dark-matter properties and early star formation.** Projected gas distribution in CDM (a) and warm dark matter (WDM; b) simulations at  $z = 20$ . If the power in the primordial density spectrum is reduced on small scales, the first stars will form much later than in the standard CDM-based scenario. If the dark matter is warm, having a substantial velocity dispersion, density perturbations on small length scales are smoothed. The hierarchy of structure formation is then truncated at a corresponding mass scale, and the first cosmological objects could be more massive than  $10^6 M_{\odot}$ . For the case of light WDM<sup>98</sup>, gas collapses into filaments, which might then fragment into multiple stellar cores. The abundance of star-forming haloes is significantly reduced in this model. Reproduced by permission of the AAS (from ref. 99).





**Figure 4 | Radiative feedback around the first stars.** Ionized bubbles are shown in blue, and regions of high molecule abundance in green. The large residual free electron fraction inside the relic H II regions, left behind after the central star has died, rapidly catalyses the reformation of molecules. The abundance of HD molecules allows the primordial gas to cool to the temperature of the CMB, possibly leading to the formation of population III.2 stars after these regions have re-collapsed so that gas densities are sufficiently high again for gravitational instability to occur<sup>77</sup>. The latter process takes of the order of the local Hubble time, thus imposing a  $\sim 100$  Myr delay in star formation. The relatively high molecule abundance in relic H II regions, along with their increasing volume-filling fraction, leads to a large optical depth to Lyman–Werner photons over physical distances of the order of several kiloparsecs (ref. 47). The development of a high optical depth to Lyman–Werner photons over such short length-scales, combined with a rapidly increasing volume filling fraction of relic H II regions, suggests that the optical depth to Lyman–Werner photons over cosmological scales may be very high, acting to suppress the build-up of a background Lyman–Werner radiation field, and mitigating negative feedback on star formation<sup>75</sup>. Note the strongly clustered nature of early star formation. Visualization courtesy of the Texas Advanced Computing Center (based on data from ref. 47).

elements. Population III stars in the range  $140\text{--}260 M_{\odot}$  explode as pair-instability supernovae, which disrupt the entire progenitor, with explosion energies ranging from  $10^{51}$  erg to  $10^{53}$  erg, and nucleosynthetic yields, defined as the heavy-element mass fraction, up to 0.5. Such supernovae exhibit an odd-even effect in the nuclei produced that is much greater than observed in any star to date, and as a result they cannot make a significant contribution to the metals observed in very low-metallicity stars today<sup>57</sup>. On the other hand, the pair-instability supernova signature may exist in a tiny fraction of the stars with intermediate metallicity ( $\sim 0.01 Z_{\odot}$ , where  $Z_{\odot}$  indicates solar metallicity), because the enrichment from even a single pair-instability supernova already endows the surrounding material with heavy elements to levels that are above the regime typically probed by surveys of metal-poor stars<sup>58</sup>.

The first stars may have been born rapidly rotating, however, and rotation can entirely modify these results<sup>59</sup>. For sufficiently high rotation rates, rotationally induced mixing is able to render the cores chemically homogeneous; mixing of heavy elements to the surface in the late stages of evolution can lead to substantial mass loss. If the cores maintain a sufficiently high rotation at the time of the supernova, it is possible to produce a long  $\gamma$ -ray burst or a jet-induced energetic supernova/hypernova<sup>60,61</sup>, with significant effects on the abundances of the ejected metals<sup>62</sup>. Large uncertainties remain in the evolutionary calculations owing to the effects of dynamo-generated magnetic fields.

The strong mechanical and chemical feedback effects exerted by explosions of population III stars have been investigated with a number of detailed calculations<sup>63–69</sup>. The key question is how the initially metal-free Universe was enriched with the first heavy chemical elements<sup>70</sup>. Recently, it has become feasible to address this process with

realistic three-dimensional simulations that start from cosmological initial conditions, and that resolve the detailed physics of the supernova blast wave expansion<sup>63,64</sup>. These simulations have shown that early enrichment is very inhomogeneous, as the low-density voids are enriched before any metals can reach into the denser filaments and virialized haloes<sup>71</sup>.

### Assembly of the first galaxies

The characteristic mass of the first star formation sites has been determined to be  $\sim 10^6 M_{\odot}$  (refs 14, 72), whereas the critical mass for hosting the formation of the first galaxies is still not known with any certainty. A promising theoretical *Ansatz* is to explore atomic cooling haloes—with  $\sim 10^8 M_{\odot}$  and virial temperatures greater than  $\sim 10^4$  K so that atomic line cooling is efficient—as their formation sites<sup>73,74</sup>. The simulations, starting from cosmological initial conditions, are just now approaching the resolution and physical realism required to investigate whether atomic cooling haloes fulfil the criteria for a first galaxy as defined above. Quite generically, in such models, the first generation of stars forms before galaxies do, and feedback effects from the first stars are expected to play a key role in determining the initial conditions for the formation of the first galaxies. Although substantial uncertainties in the overall formation efficiency of the first stars still remain, it is possible, and perhaps probable, that at least one primordial star had formed in the region that was destined to eventually become a first galaxy<sup>75</sup>. If the early generation stars were massive,  $\gtrsim 10 M_{\odot}$ , the feedback effects described in the previous section would shape the conditions for subsequent star formation in the region.

The gas expelled by the H II regions and supernovae of the first stars would have been too hot and diffuse to allow further star formation until it had time to cool, as well as to reach high densities again in the course of being reincorporated in a growing dark-matter halo. Both cooling and re-collapse occur rather slowly, thus rendering star formation intermittent in the early formation phase of the first galaxies. Analytic models<sup>76</sup> and detailed numerical simulations<sup>47,77</sup> both show that the gas re-incorporation time is as long as  $10^8$  years, roughly corresponding to the dynamical time for a first-galaxy halo to be assembled.

Chemical enrichment by the first supernovae is among the most important processes in the formation of the first galaxies. Efficient cooling by metal lines and dust thermal emission regulate the temperature of already metal-enriched population II (see Box 1) star-forming regions in the first galaxies. The concept of a ‘critical metallicity’ has been introduced to characterize the transition of the star-formation mode from predominantly high-mass, population III or population II, to low-mass population II stars<sup>78</sup>. However, this critical gas metallicity is still poorly determined. It is not even clear if there exists such a sharp transition. Some studies show that even a slight quantity of metals in a gas may be enough to change the gas thermal evolution significantly<sup>79</sup>, whereas others argue that the cooling efficiency at low densities<sup>80</sup> is crucial and is significantly enhanced only above  $10^{-4} Z_{\odot}$ . As the enrichment from even a single pair-instability supernova by a very massive population III star probably leads to metallicities of  $Z > 10^{-2} Z_{\odot}$  (ref. 63), well in excess of any predicted value for the critical metallicity, these arguments might be somewhat academic. The characteristic mass of pre-stellar gas clumps is probably determined by a number of physical processes (for example, turbulence and, possibly, dynamo-amplified primordial magnetic fields) other than radiative cooling. The overall effect of gas metallicity on star formation may well be limited<sup>81</sup>.

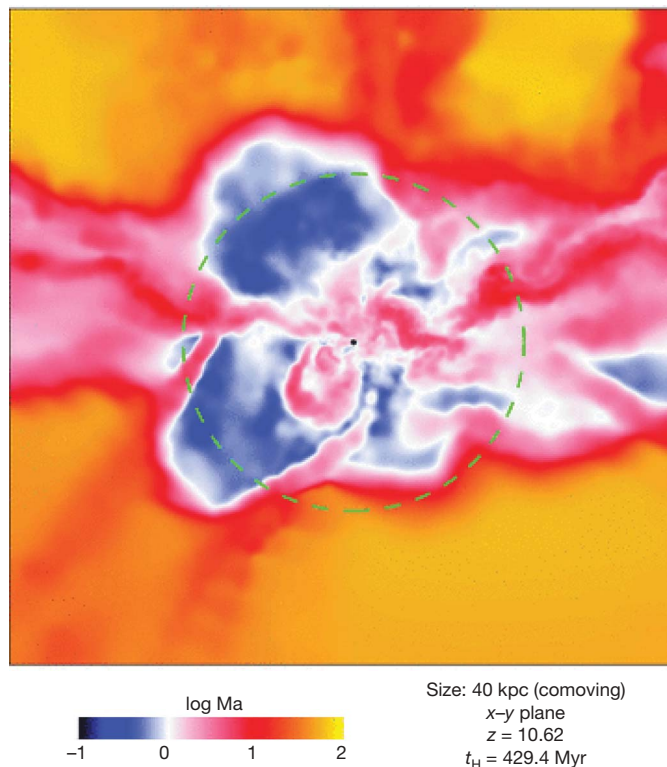
Recent cosmological simulations have demonstrated that star formation inside the first galaxies was strongly influenced by gravitationally driven supersonic turbulence that was generated during the virialization process<sup>64,73,74</sup>. This is in marked contrast to the rather quiescent, quasi-hydrostatic situation in minihaloes (see Fig. 5). It thus appears possible that the first galaxies harboured the first stellar clusters, if present-day star formation offers any guide here, where it

is widely believed that gravo-turbulent fragmentation is responsible for shaping the initial mass function<sup>24,82</sup>. It is an open question as to whether the first galaxies could have harboured the first globular clusters, which are the oldest star clusters known.

### Future empirical probes

Studying the formation of the first stars and galaxies will be at the frontier of astronomy and cosmology in the next decade. Astronomers will muster a comprehensive arsenal of observational probes. The most prominent among these concern the optical depth to Thomson scattering of cosmic microwave background photons off free electrons<sup>83–85</sup>, the near-infrared background<sup>86</sup>, high-redshift  $\gamma$ -ray bursts<sup>87–89</sup>, the possibility of scrutinizing the nature of the first stars by metals found in the oldest Galactic halo stars, dubbed ‘stellar archaeology’<sup>90,91</sup>, and various facilities now being deployed to map reionization using the redshifted 21 cm line of neutral hydrogen<sup>92–94</sup>. The James Webb Space Telescope (JWST) will perform a number of observations designed to test key assumptions of our current theory of the first stars and galaxies<sup>95</sup>. How could the existence of massive population III stars be unambiguously inferred? The most clear-cut diagnostic is the ratio of recombination lines emitted from the H II regions around single population III stars, or clusters thereof, to be measured with ultra-deep near-infrared and mid-infrared spectroscopy. Owing to the high effective temperature of the population III stellar continuum,  $\sim 10^5$  K, strong He II line emission at a rest-frame wavelength of 1,640 Å is predicted, with a ratio compared to Lyman- $\alpha$  that is one to two orders of magnitude larger than for normal stars<sup>41</sup>. A second crucial observational campaign aims at a census of very high- $z$  supernovae<sup>96</sup> through deep broadband near-infrared imaging. One key objective is to search for possible pair-instability

supernova events, which would clearly stand out owing to their extreme intrinsic brightness, as well as their very long durations—a few years in the observer frame<sup>97</sup>. The goal of making useful predictions for the high-redshift frontier is now clearly moving within reach, and the pace of progress is likely to be rapid.



**Figure 5 | Turbulence inside the first galaxies.** Shown is the Mach number (Ma) in a slice through the central 40 kpc (comoving) of the galaxy. The dashed line denotes the virial radius of  $\sim 1$  kpc. The Mach number approaches unity at the virial shock, where the accreted gas is heated to the virial temperature. Inflows of cold gas along filaments are supersonic by a factor of  $\sim 10$ , resulting in strong turbulent flows in the galactic centre. The age of the Universe at redshift  $z \approx 10$  is given by  $t_H$ . Reproduced by permission of Wiley-Blackwell (from ref. 74).

1. Barkana, R. & Loeb, A. In the beginning: The first sources of light and the reionization of the Universe. *Phys. Rep.* **349**, 125–238 (2001).
2. Miralda-Escudé, J. The dark age of the Universe. *Science* **300**, 1904–1909 (2003).
3. Loeb, A., Ferrara, A. & Ellis, R. S. *First Light in the Universe* (Springer, 2008).
4. Bromm, V. & Larson, R. B. The first stars. *Annu. Rev. Astron. Astrophys.* **42**, 79–118 (2004).
5. Ciardi, B. & Ferrara, A. The first cosmic structures and their effects. *Space Sci. Rev.* **116**, 625–705 (2005).
6. Blumenthal, G. R., Faber, S. M., Primack, J. R. & Rees, M. J. Formation of galaxies and large-scale structure with cold dark matter. *Nature* **311**, 517–525 (1984).
7. Couchman, H. M. P. & Rees, M. J. Pregalactic evolution in cosmologies with cold dark matter. *Mon. Not. R. Astron. Soc.* **221**, 53–62 (1986).
8. Glover, S. The formation of the first stars in the Universe. *Space Sci. Rev.* **117**, 445–508 (2005).
9. Dekel, A. et al. Cold streams in early massive hot haloes as the main mode of galaxy formation. *Nature* **457**, 451–454 (2009).
10. Galli, D. & Palla, F. The chemistry of the early Universe. *Astron. Astrophys.* **335**, 403–420 (1998).
11. Abel, T., Bryan, G. L. & Norman, M. L. The formation of the first star in the Universe. *Science* **295**, 93–98 (2002).
12. Bromm, V., Coppi, P. S. & Larson, R. B. The formation of the first stars. I. The primordial star-forming cloud. *Astrophys. J.* **564**, 23–51 (2002).
13. Yoshida, N., Omukai, K., Hernquist, L. & Abel, T. Formation of primordial stars in a  $\Lambda$ CDM Universe. *Astrophys. J.* **652**, 6–25 (2006).
14. Tegmark, M. et al. How small were the first cosmological objects? *Astrophys. J.* **474**, 1–12 (1997).
15. Gao, L. et al. The first generation of stars in the  $\Lambda$  cold dark matter cosmology. *Mon. Not. R. Astron. Soc.* **378**, 449–468 (2007).
16. Omukai, K. & Nishi, R. Photodissociative regulation of star formation in metal-free pregalactic clouds. *Astrophys. J.* **518**, 64–68 (1999).
17. Kitayama, T., Yoshida, N., Susa, H. & Umemura, M. The structure and evolution of early cosmological H II regions. *Astrophys. J.* **613**, 631–645 (2004).
18. Machida, M. N., Omukai, K., Matsumoto, T. & Inutsuka, S. Conditions for the formation of first-stars binaries. *Astrophys. J.* **677**, 813–827 (2008).
19. Clark, P. C., Glover, S. C. O. & Klessen, R. S. The first stellar cluster. *Astrophys. J.* **672**, 757–764 (2008).
20. Yoshida, N., Omukai, K. & Hernquist, L. Protostar formation in the early Universe. *Science* **321**, 669–671 (2008).
21. Silk, J. The first stars. *Mon. Not. R. Astron. Soc.* **205**, 705–718 (1983).
22. Bromm, V. & Loeb, A. Accretion onto a primordial protostar. *N. Astron.* **9**, 353–364 (2004).
23. Bromm, V., Coppi, P. S. & Larson, R. B. Forming the first stars in the Universe: The fragmentation of primordial gas. *Astrophys. J.* **527**, L5–L9 (1999).
24. McKee, C. F. & Ostriker, E. C. Theory of star formation. *Annu. Rev. Astron. Astrophys.* **45**, 565–687 (2007).
25. McKee, C. F. & Tan, J. C. The formation of the first stars II. Radiative feedback processes and implications for the initial mass function. *Astrophys. J.* **681**, 771–797 (2008).
26. Ryu, D., Kang, H., Cho, J. & Das, S. Turbulence and magnetic fields in the large-scale structure of the Universe. *Science* **320**, 909–912 (2008).
27. Schaerer, D. On the properties of massive population III stars and metal-free stellar populations. *Astron. Astrophys.* **382**, 28–42 (2002).
28. McKee, C. F. & Tan, J. C. Massive star formation in 100,000 years from turbulent and pressurised molecular clouds. *Nature* **416**, 59–61 (2002).
29. Tan, J. C. & McKee, C. F. The formation of the first stars. I. Mass infall rates, accretion disk structure, and protostellar evolution. *Astrophys. J.* **603**, 383–400 (2004).
30. Hollenbach, D. J., Johnstone, D., Lizano, S. & Shu, F. Photoevaporation of disks around massive stars and application to ultracompact H II regions. *Astrophys. J.* **428**, 654–669 (1994).
31. Tan, J. C. & Blackman, E. G. Protostellar disk dynamos and hydromagnetic outflows in primordial star formation. *Astrophys. J.* **603**, 401–413 (2003).
32. Stacy, A. & Bromm, V. Impact of cosmic rays on Population III star formation. *Mon. Not. R. Astron. Soc.* **382**, 229–238 (2007).



33. Yoshida, N., Omukai, K. & Hernquist, L. Formation of massive primordial stars in a reionized gas. *Astrophys. J.* **667**, L117–L120 (2007).
34. Nakamura, F. & Umemura, M. The stellar initial mass function in primordial galaxies. *Astrophys. J.* **569**, 549–557 (2002).
35. Johnson, J. L. & Bromm, V. The cooling of shock-compressed primordial gas. *Mon. Not. R. Astron. Soc.* **366**, 247–256 (2006).
36. McGreer, I. D. & Bryan, G. L. The impact of HD cooling on the formation of the first stars. *Astrophys. J.* **685**, 8–20 (2008).
37. Spolyar, D., Freese, K. & Gondolo, P. Dark matter and the first stars: A new phase of stellar evolution. *Phys. Rev. Lett.* **100**, 051101 (2008).
38. Freese, K., Bodenheimer, P., Spolyar, D. & Gondolo, P. Stellar structure of dark stars: A first phase of stellar evolution resulting from dark matter annihilation. *Astrophys. J.* **685**, L101–L104 (2008).
39. Iocco, F. *et al.* Dark matter annihilation effects on the first stars. *Mon. Not. R. Astron. Soc.* **390**, 1655–1669 (2008).
40. Griest, K. & Kamionkowski, M. Supersymmetric dark matter. *Phys. Rep.* **333**, 167–182 (2000).
41. Bromm, V., Kudritzki, R. P. & Loeb, A. Generic spectrum and ionization efficiency of a heavy initial mass function for the first stars. *Astrophys. J.* **552**, 464–472 (2001).
42. Whalen, D., Abel, T. & Norman, M. L. Radiation hydrodynamic evolution of primordial H II regions. *Astrophys. J.* **610**, 14–22 (2004).
43. Alvarez, M. A., Bromm, V. & Shapiro, P. R. The H II region of the first star. *Astrophys. J.* **639**, 621–632 (2006).
44. Abel, T., Wise, J. H. & Bryan, G. L. The H II region of a primordial star. *Astrophys. J.* **659**, L87–L90 (2007).
45. Ciardi, B., Ferrara, A. & Abel, T. Intergalactic H<sub>2</sub> photodissociation and the soft ultraviolet background produced by population III objects. *Astrophys. J.* **533**, 594–600 (2000).
46. Haiman, Z., Abel, T. & Rees, M. J. The radiative feedback of the first cosmological objects. *Astrophys. J.* **534**, 11–24 (2000).
47. Johnson, J. L., Greif, T. H. & Bromm, V. Radiative feedback in the formation of the first protogalaxies. *Astrophys. J.* **665**, 85–95 (2007).
48. Susa, H. & Umemura, M. Secondary star formation in a population III object. *Astrophys. J.* **645**, L93–L96 (2006).
49. Ahn, K. & Shapiro, P. R. Does radiative feedback by the first stars promote or prevent second generation star formation? *Mon. Not. R. Astron. Soc.* **375**, 881–908 (2007).
50. Whalen, D., O'Shea, B. W., Smidt, J. & Norman, M. L. How the first stars regulated local star formation. I. Radiative feedback. *Astrophys. J.* **679**, 925–941 (2008).
51. Machacek, M. E., Bryan, G. L. & Abel, T. Simulations of pregalactic structure formation with radiative feedback. *Astrophys. J.* **548**, 509–521 (2001).
52. Ricotti, M., Gnedin, N. Y. & Shull, J. M. Feedback from galaxy formation: Production and photodissociation of primordial H<sub>2</sub>. *Astrophys. J.* **560**, 580–591 (2001).
53. Oh, S. P. & Haiman, Z. Second-generation objects in the Universe: Radiative cooling and collapse of halos with virial temperatures above 10<sup>4</sup> K. *Astrophys. J.* **569**, 558–572 (2002).
54. Ferrara, A. The positive feedback of population III objects on galaxy formation. *Astrophys. J.* **499**, L17–L20 (1998).
55. Haiman, Z., Rees, M. J. & Loeb, A. Destruction of molecular hydrogen during cosmological reionization. *Astrophys. J.* **476**, 458–463 (1997).  
**Introduced the concept of negative radiative feedback that might act to self-limit population III star formation.**
56. Heger, A. & Woosley, S. E. The nucleosynthetic signature of population III. *Astrophys. J.* **567**, 532–543 (2002).  
**The first definitive calculation of the nucleosynthetic pattern of very massive population III stars that die as pair-instability supernovae.**
57. Tumlinson, J. Chemical evolution in hierarchical models of cosmic structure. I. Constraints on the early stellar initial mass function. *Astrophys. J.* **641**, 1–20 (2006).
58. Karlsson, T., Johnson, J. L. & Bromm, V. Uncovering the signature of the first stars in the Universe. *Astrophys. J.* **679**, 6–16 (2008).
59. Maeder, A., Meynet, G. & Ekström, S. in *From Stars to Galaxies: Building the Pieces to Build Up the Universe* (eds Vallenari, A., Tantaló, R., Portinari, L. & Moretti, A.) 13–20 (ASP Conf. Ser. 374, Astronomical Society of the Pacific, 2007).
60. Yoon, S.-C. & Langer, N. Evolution of rapidly rotating metal-poor massive stars towards gamma-ray bursts. *Astron. Astrophys.* **443**, 643–648 (2005).
61. Woosley, S. E. & Heger, A. The progenitor stars of gamma-ray bursts. *Astrophys. J.* **637**, 914–921 (2006).
62. Nomoto, K., Tominaga, N., Tanaka, M., Maeda, K. & Umeda, H. in *Massive stars as Cosmic Engines* (eds Bresolin, F., Crowther, P. A. & Puls, J.) 463–470 (Cambridge Univ. Press, 2008).
63. Greif, T. H., Johnson, J. L., Bromm, V. & Klessen, R. S. The first supernova explosions: Energetics, feedback, and chemical enrichment. *Astrophys. J.* **670**, 1–14 (2007).
64. Wise, J. H. & Abel, T. Resolving the formation of protogalaxies. III. Feedback from the first stars. *Astrophys. J.* **685**, 40–56 (2008).
65. Mori, M., Ferrara, A. & Madau, P. Early metal enrichment by pregalactic outflows. II. Three-dimensional simulations of blow-away. *Astrophys. J.* **571**, 40–55 (2002).
66. Bromm, V., Yoshida, N. & Hernquist, L. The first supernova explosions in the Universe. *Astrophys. J.* **596**, L135–L138 (2003).
67. Kitayama, T. & Yoshida, N. Supernova explosions in the early Universe: Evolution of radiative remnants and the halo destruction efficiency. *Astrophys. J.* **630**, 675–688 (2005).
68. Machida, M. N., Tomisaka, K., Nakamura, F. & Fujimoto, M. Y. Low-mass star formation triggered by supernovae in primordial clouds. *Astrophys. J.* **622**, 39–57 (2005).
69. Whalen, D., van Veelen, B., O'Shea, B. W. & Norman, M. L. The destruction of cosmological minihalos by primordial supernovae. *Astrophys. J.* **682**, 49–67 (2008).
70. Tornatore, L., Ferrara, A. & Schneider, R. Population III stars: Hidden or disappeared? *Mon. Not. R. Astron. Soc.* **382**, 945–950 (2007).
71. Cen, R. & Riquelme, M. A. Lower metal enrichment of virialized gas in minihalos. *Astrophys. J.* **674**, 644–652 (2008).
72. Yoshida, N., Abel, T., Hernquist, L. & Sugiyama, N. Simulations of early structure formation: Primordial gas clouds. *Astrophys. J.* **592**, 645–663 (2003).
73. Wise, J. H. & Abel, T. Resolving the formation of protogalaxies. I. Virialization. *Astrophys. J.* **665**, 899–910 (2007).
74. Greif, T. H., Johnson, J. L., Klessen, R. S. & Bromm, V. The first galaxies: Assembly, cooling and the onset of turbulence. *Mon. Not. R. Astron. Soc.* **387**, 1021–1036 (2008).
75. Johnson, J. L., Greif, T. H. & Bromm, V. Occurrence of metal-free galaxies in the early Universe. *Mon. Not. R. Astron. Soc.* **388**, 26–38 (2008).
76. Yoshida, N., Bromm, V. & Hernquist, L. The era of massive population III stars: Cosmological implications and self-termination. *Astrophys. J.* **605**, 579–590 (2004).
77. Yoshida, N., Oh, S.-P., Kitayama, T. & Hernquist, L. Early cosmological H II/He III regions and their impact on second-generation star formation. *Astrophys. J.* **663**, 687–707 (2007).
78. Bromm, V., Ferrara, A., Coppi, P. S. & Larson, R. B. The fragmentation of pre-enriched primordial objects. *Mon. Not. R. Astron. Soc.* **328**, 969–976 (2001).
79. Omukai, K., Tsuribe, T., Ferrara, A. & Schneider, R. Thermal and fragmentation properties of star-forming clouds in low-metallicity environments. *Astrophys. J.* **626**, 627–643 (2005).
80. Bromm, V. & Loeb, A. The fragmentation of pre-enriched primordial objects. *Nature* **425**, 812–814 (2003).
81. Jappsen, A.-K., Klessen, R. S., Glover, S. C. O. & MacLow, M.-M. Star formation at very low metallicity. IV. Fragmentation does not depend on metallicity for cold initial conditions. *Astrophys. J.* (in the press); preprint at (<http://arXiv.org/abs/0709.3530>) (2007).
82. Mac Low, M.-M. & Klessen, R. S. Control of star formation by supersonic turbulence. *Rev. Mod. Phys.* **76**, 125–194 (2004).
83. Fan, X., Carilli, C. L. & Keating, B. Observational constraints on cosmic reionization. *Annu. Rev. Astron. Astrophys.* **44**, 415–462 (2006).
84. Komatsu, E. *et al.* Five-year Wilkinson Microwave Anisotropy Probe observations: Cosmological interpretation. *Astrophys. J.* **180** (Suppl.), 330–376 (2009).
85. Holder, G. P., Haiman, Z., Kaplinghat, M. & Knox, L. The reionization history at high redshifts. II. Estimating the optical depth to Thomson scattering from cosmic microwave background polarization. *Astrophys. J.* **595**, 13–18 (2003).
86. Kashlinsky, A., Arendt, R. G., Mather, J. & Moseley, S. H. Tracing the first stars with fluctuations of the cosmic infrared background. *Nature* **438**, 45–50 (2005).
87. Lamb, D. Q. & Reichardt, D. E. Gamma-ray bursts as a probe of the very high redshift Universe. *Astrophys. J.* **536**, 1–18 (2000).
88. Bromm, V. & Loeb, A. High-redshift gamma-ray bursts from population III progenitors. *Astrophys. J.* **642**, 382–388 (2006).
89. Totani, T. *et al.* Implications for cosmic reionization from the optical afterglow spectrum of the gamma-ray burst 050904 at  $z = 6.3$ . *Publ. Astron. Soc. Jpn.* **58**, 485–498 (2006).
90. Frebel, A., Johnson, J. L. & Bromm, V. Probing the formation of the first low-mass stars with stellar archaeology. *Mon. Not. R. Astron. Soc.* **380**, L40–L44 (2007).
91. Iwamoto, N., Umeda, H., Tominaga, N., Nomoto, K. & Maeda, K. The first chemical enrichment in the Universe and the formation of hyper metal-poor stars. *Science* **309**, 451–453 (2005).
92. Furlanetto, S. R., Oh, S. P. & Briggs, F. H. Cosmology at low frequencies: The 21 cm transition and the high-redshift Universe. *Phys. Rep.* **433**, 181–301 (2006).
93. Madau, P., Meiksin, A. & Rees, M. J. 21 centimeter tomography of the intergalactic medium at high redshift. *Astrophys. J.* **475**, 429–444 (1997).
94. McQuinn, M. *et al.* The morphology of H II regions during reionization. *Mon. Not. R. Astron. Soc.* **377**, 1043–1063 (2007).
95. Gardner, J. P. *et al.* The James Webb Space Telescope. *Space Sci. Rev.* **123**, 485–606 (2006).
96. Mackey, J., Bromm, V. & Hernquist, L. Three epochs of star formation in the high-redshift Universe. *Astrophys. J.* **586**, 1–11 (2003).
97. Scannapieco, E., Madau, P., Woosley, S. E., Heger, A. & Ferrara, A. The detectability of pair-production supernovae at  $z = 6$ . *Astrophys. J.* **633**, 1031–1041 (2005).
98. Gao, L. & Theuns, T. Lighting the Universe with filaments. *Science* **317**, 1527–1530 (2007).
99. Yoshida, N., Sokasian, A., Hernquist, L. & Springel, V. Early structure formation and reionization in a warm dark matter cosmology. *Astrophys. J.* **591**, L1–L4 (2003).
100. Beers, T. C. & Christlieb, N. The discovery and analysis of very metal-poor stars in the Galaxy. *Annu. Rev. Astron. Astrophys.* **43**, 531–580 (2005).

**Acknowledgements** We are grateful for the hospitality of KITP, University of California Santa Barbara. This work was supported in part by NSF and NASA.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence should be addressed to V.B. ([vbromm@astro.as.utexas.edu](mailto:vbromm@astro.as.utexas.edu)).

# HDAC2 negatively regulates memory formation and synaptic plasticity

Ji-Song Guan<sup>1,2,3\*</sup>, Stephen J. Haggarty<sup>3,4\*</sup>, Emanuela Giacometti<sup>5,6\*</sup>, Jan-Hermen Dannenberg<sup>7\*†</sup>, Nadine Joseph<sup>1,2,3</sup>, Jun Gao<sup>1,2</sup>, Thomas J. F. Nieland<sup>3</sup>, Ying Zhou<sup>1,2</sup>, Xinyu Wang<sup>1,2</sup>, Ralph Mazitschek<sup>3,8</sup>, James E. Bradner<sup>3</sup>, Ronald A. DePinho<sup>7</sup>, Rudolf Jaenisch<sup>5,6</sup> & Li-Huei Tsai<sup>1,2,3</sup>

Chromatin modifications, especially histone-tail acetylation, have been implicated in memory formation. Increased histone-tail acetylation induced by inhibitors of histone deacetylases (HDACis) facilitates learning and memory in wild-type mice as well as in mouse models of neurodegeneration. Harnessing the therapeutic potential of HDACis requires knowledge of the specific HDAC family member(s) linked to cognitive enhancement. Here we show that neuron-specific overexpression of HDAC2, but not that of HDAC1, decreased dendritic spine density, synapse number, synaptic plasticity and memory formation. Conversely, *Hdac2* deficiency resulted in increased synapse number and memory facilitation, similar to chronic treatment with HDACis in mice. Notably, reduced synapse number and learning impairment of HDAC2-overexpressing mice were ameliorated by chronic treatment with HDACis. Correspondingly, treatment with HDACis failed to further facilitate memory formation in *Hdac2*-deficient mice. Furthermore, analysis of promoter occupancy revealed an association of HDAC2 with the promoters of genes implicated in synaptic plasticity and memory formation. Taken together, our results suggest that HDAC2 functions in modulating synaptic plasticity and long-lasting changes of neural circuits, which in turn negatively regulates learning and memory. These observations encourage the development and testing of HDAC2-selective inhibitors for human diseases associated with memory impairment.

Neuronal adaptive responses implicated in memory formation and storage involve functional and structural synaptic changes, which require alterations in gene expression<sup>1,2</sup>. The mechanisms underlying this process are still unclear. Chromatin remodelling, especially through histone-tail acetylation, which alters the compact chromatin structure and changes the accessibility of DNA to regulatory proteins, is emerging as a fundamental mechanism for regulating gene expression<sup>3,4</sup>. Recently, histone acetylation has been implicated in synaptic plasticity and learning behaviour<sup>5–9</sup>. In particular, it was shown that the non-selective HDAC inhibitor sodium butyrate reinstated learning ability and promoted the retrieval of long-term memory in mice even after massive neuronal loss<sup>10</sup>. Taken together, these observations indicate that HDAC inhibition may provide a therapeutic avenue for memory impairment caused by neurodegenerative and other diseases. However, the clinical application of broadly acting non-selective HDACis in cancer has shown a range of untoward effects, which may limit their utility in application to non-lethal, chronic diseases<sup>11–13</sup>. Thus, identification of the HDAC family member(s) specifically involved in memory formation will help in discovering the mechanism(s) by which chromatin remodelling regulates memory and lead to more selective HDACis for memory enhancement.

## Chronic HDACi treatment facilitates memory formation

On the basis of HDAC selectivity studies *in vitro*, sodium butyrate might affect brain function mainly through class I HDACs, including *Hdac1*, *Hdac2*, *Hdac3* and *Hdac8* (Supplementary Information). To

gain further insight into the ability of systemic HDACis to enhance memory, and to identify more potent HDACis than sodium butyrate<sup>10</sup>, a set of prototypical HDACis with a wide range of structures and metal-chelating elements were tested in mice using a contextual fear conditioning assay. While some HDACis did not facilitate memory formation (data not shown), chronic treatment with suberoylanilide hydroxamic acid (SAHA; vorinostat), a clinically approved agent, enhanced memory formation more potently than sodium butyrate (Supplementary Figs 1 and 16c, d). This finding is consistent with the previous report<sup>3</sup> that SAHA rescued memory deficits in a mouse model of Rubinstein–Taybi syndrome. Given the ability of both compounds to cross the blood–brain barrier, as measured indirectly by changes in histone acetylation in brain<sup>10,14,15</sup> and directly by measuring the amount of SAHA in the brain (data not shown), we focused our subsequent studies on SAHA-induced memory enhancement.

Because SAHA inhibits primarily class I HDACs and HDAC6 (ref. 16), a class II HDAC known to target acetylation of  $\alpha$ -tubulin on Lys40 ( $\alpha$ -tubulin(K40))<sup>17,18</sup>, we examined whether selective inhibition of HDAC6 with the HDACi WT-161 had effects on memory formation (Supplementary Fig. 2a–c). WT-161 did not increase memory formation (Supplementary Fig. 2d), suggesting that HDAC6 inhibition is not responsible for SAHA-induced memory enhancement. In agreement with these observations, proteome-wide studies of a SAHA-based affinity probe identified HDAC1 and HDAC2 as the main cellular targets<sup>19</sup>, indicating that these may be relevant targets for HDACi-induced memory enhancement.

<sup>1</sup>Picower Institute for Learning and Memory, Department of Brain and Cognitive Sciences, <sup>2</sup>Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. <sup>3</sup>Stanley Center for Psychiatric Research, Broad Institute of Harvard University and Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA. <sup>4</sup>Center for Human Genetic Research, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts 02142, USA. <sup>5</sup>Whitehead Institute for Biomedical Research, Cambridge, Massachusetts 02142, USA. <sup>6</sup>Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. <sup>7</sup>Belfer Institute for Applied Cancer Science, Departments of Medical Oncology, Medicine and Genetics, Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>8</sup>Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 240 Longwood Avenue, Boston, Massachusetts 02115, USA. <sup>†</sup>Present address: Division of Molecular Genetics, Netherlands Cancer Institute, Amsterdam 1066 CX, The Netherlands.

\*These authors contributed equally to this work.



### Hdac2, but not Hdac1, functions in memory formation

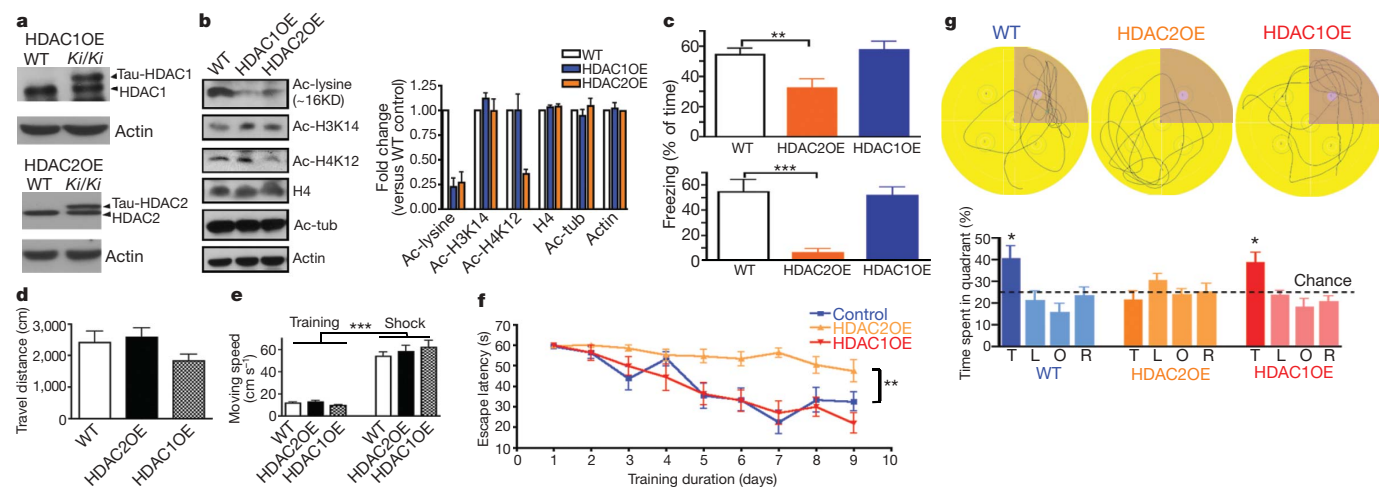
To directly evaluate the physiological role of HDAC1 and HDAC2 in the brain, we generated two mouse lines in which HDAC1 or HDAC2 was overexpressed in neurons (Supplementary Fig. 3a–c). The mouse *Hdac1* or *Hdac2* coding sequence was placed in frame with the endogenous initiation codon of exon 1 of the *Tau* (also known as *Mapt*) gene, thereby creating a fusion protein containing the first 31 amino-acid residues of *Tau*. Previously, homozygous animals mutant for *Tau* were shown to be indistinguishable from wild-type littermates in memory tests<sup>20,21</sup>. A two- to threefold increase in HDAC1 or HDAC2 protein expression in brain of homozygous animals in comparison with wild-type mice was observed in the hippocampus and other areas of the brain (Fig. 1a and Supplementary Fig. 4). Consistently, the overall level of acetylated lysine was decreased in homozygous HDAC1-overexpressing (HDAC1OE) and HDAC2-overexpressing (HDAC2OE) mice (Fig. 1b), especially in the pyramidal neurons of the hippocampal formation (Supplementary Fig. 3d). We found decreased acetylation of histone 4 on Lys 12 (H4K12) and H4K5, but not of H3K14, in HDAC2OE mouse brains (Fig. 1b and data not shown). In contrast, the level of acetylated  $\alpha$ -tubulin(K40) did not change in HDAC1OE or HDAC2OE mice. Thus, the HDAC1/2-overexpressing animals showed increased histone deacetylation in the brain in comparison with wild-type littermates. There was no discernible difference in gross brain anatomy or neuronal positioning in the HDAC1/2-overexpressing mice (Supplementary Figs 5 and 12), suggesting that increased HDAC1/2 is not detrimental to brain development.

To evaluate the consequence of HDAC1/2 overexpression on hippocampus-dependent memory formation, four-month-old HDAC1OE and HDAC2OE mice were trained using Pavlovian fear conditioning paradigm before a memory test 24 h later. Whereas HDAC1OE mice ( $n = 17$ ) displayed a comparable freezing level ( $P = 0.6470$ , context;  $P = 0.8199$ , tone) to that of the control mice ( $n = 19$ ), we found that HDAC2OE mice ( $n = 14$ ) showed markedly decreased freezing behaviour in both context-dependent and tone-dependent fear learning (Fig. 1c,  $P = 0.0035$  and  $P = 0.0007$ , respectively). The decrease in freezing behaviour in HDAC2OE mice was not due to motor defects or impaired pain sensation, because their response to electric foot shock and the explorative behaviour during

exposure to novel context were similar to those of the control mice (Fig. 1d, e). In the short-term memory test, no significant difference could be detected between HDAC1OE, HDAC2OE and wild-type control mice in both context-dependent and tone-dependent fear learning 3 h after training (Supplementary Fig. 6b). These observations suggest that a gain of HDAC2 function (but not that of HDAC1) in the nervous system results in associative learning impairments.

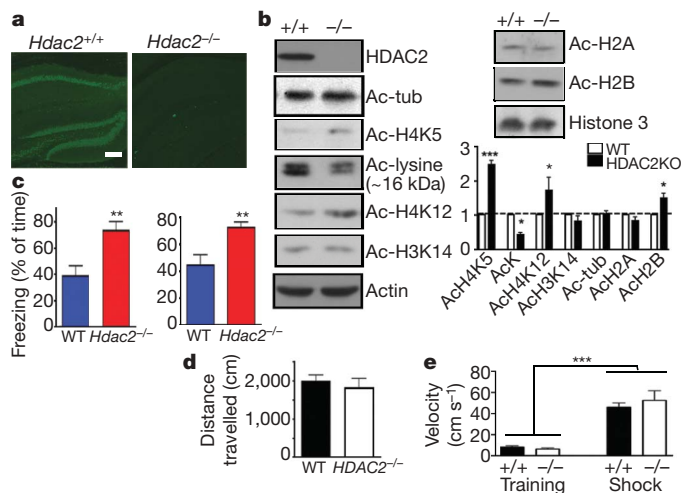
To evaluate the integrity of hippocampus-dependent memory formation in the HDAC2OE mice further, we used the Morris water maze paradigm. The escape latency of the HDAC1OE mice closely paralleled that of the control mice during a 9-day training period for the hidden-platform model ( $P = 0.7784$ , two-way analysis of variance (ANOVA)). In contrast, the HDAC2OE mice showed significantly increased escape latency throughout the training process ( $P = 0.0010$ , two-way ANOVA; Fig. 1f). A probe trial was also performed to quantitatively measure the time spent in each quadrant of the swimming pool when the hidden platform was removed from the pool. Notably, the HDAC2OE mice did not show a preference for the target quadrant, whereas the control mice ( $P = 0.0145$ , target quadrant versus opposite quadrant) and HDAC1OE mice ( $P = 0.008$ ) spent significantly more time in the target quadrant (Fig. 1g). Comparable motor and visual function between the various strains were observed in the visible-platform test (Supplementary Fig. 6a, b). These results revealed a marked decrease in spatial learning of the HDAC2OE mice. Furthermore, HDAC2OE (but not HDAC1OE) mice showed impairment of spatial working memory in a T-maze non-matching-to-place task (Supplementary Fig. 6h). Thus, a gain of HDAC2 function, but not that of HDAC1, impairs hippocampus-dependent memory formation as well as working memory.

To further investigate the role of HDAC2 in associative learning, *Hdac2*-deficient mice (HDAC2KO) were generated by crossing mice carrying a floxed *Hdac2* allele with *Nestin-Cre* transgenic mice. Germline deletion of *Hdac2* resulted in viable and fertile *Hdac2*<sup>+/-</sup> mice with no obvious histological abnormalities up to a year of age (Supplementary Methods and Supplementary Fig. 7). Crossing *Hdac2*<sup>+/-</sup> mice gave rise to viable *Hdac2*-deficient mice, which had normal brain anatomy and cell positioning (Fig. 2a, b, Supplementary Figs 8a and 12, and data not shown). H4K5, H4K12 and



**Figure 1 | Mice overexpressing HDAC2, but not HDAC1, show impaired memory formation.** **a**, Western blots from brain lysate showed upregulation of HDAC1 and HDAC2 in homozygous mice overexpressing HDAC1 and HDAC2, respectively. WT, wild type; *Ki/Ki*, knock-in. **b**, Decreased histone acetylation in the hippocampus of HDAC1OE and HDAC2OE mice. Tub, tubulin. **c**, Associative memory test for HDAC1OE and HDAC2OE mice (HDAC1OE,  $n = 17$ ; WT,  $n = 19$ ; HDAC2OE,  $n = 14$ ). Results are expressed as the percentage of time for which freezing behaviour was shown, during the contextual (top) and tone-dependent (bottom) memory tests. **d**, Distance travelled during the initial 3-min exposure to the training box in an open

field test. **e**, Velocity during training and electric foot shock ( $I = 1.0$  mA). **f**, Hidden platform of Morris water maze. Escape latencies of WT mice improved significantly faster than those of HDAC2OE mice (genotype  $\times$  day  $F_{1,8} = 3.401$ ,  $P = 0.0010$ ; genotype  $F_{1,8} = 52.32$ ,  $P < 0.0001$ ; day  $F_{1,8} = 7.372$ ,  $P < 0.0001$ ) but not those of HDAC1OE mice (genotype  $\times$  day  $F_{1,8} = 0.5989$ ,  $P = 0.7784$ ; day  $F_{1,8} = 12.14$ ,  $P < 0.0001$ ; genotype  $F_{1,8} = 0.2672$ ,  $P = 0.6057$ ). **g**, Representative path tracings of the probe test on day 5. Swimming times in each quadrant were measured: T, target quadrant; L, left quadrant; O, opposite quadrant; R, right quadrant. \* $P < 0.05$ ; \*\* $P < 0.005$ ; \*\*\* $P < 0.001$ . Error bars show s.e.m.



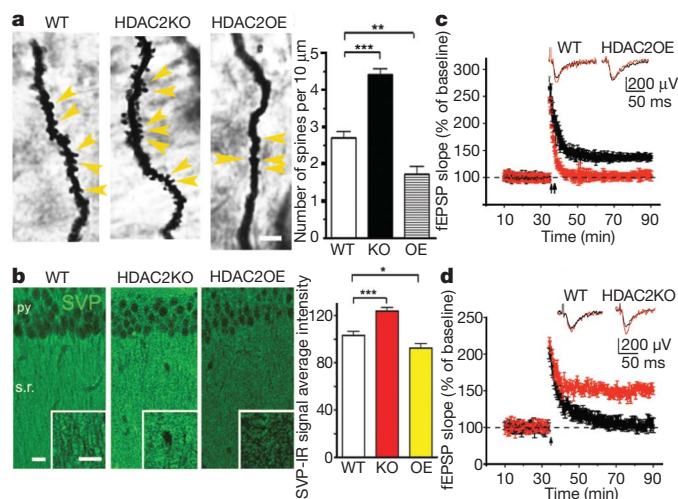
**Figure 2 | *Hdac2* knockout mice show enhanced memory formation.**

**a**, Immunohistochemical images show HDAC2 expression in the hippocampus of WT mice but not in HDAC2KO mice. Scale bar, 100  $\mu$ m. **b**, Samples from hippocampal histone extraction show changes in histone acetylation level in HDAC2KO mice. **c**, Freezing behaviour of HDAC2KO mice and control littermates (*Hdac2*<sup>-/-</sup>, *n* = 9; WT, *n* = 11) during the contextual (left) and tone-dependent (right) memory tests. **d**, Distance travelled during the initial 3-min exposure to the training box in an open field test. **e**, Velocity during training and electric foot shock (*I* = 0.8 mA). \*\**P* < 0.005. Error bars show s.e.m.

H2B acetylation were significantly increased in the hippocampus of HDAC2KO mice. However, overall lysine acetylation in histone preparations was slightly decreased, as revealed by western blot analysis (Fig. 2b). This might be the consequence of a compensatory increase in HDAC1 in HDAC2KO mice (Supplementary Fig. 7d). The HDAC2KO mice showed markedly increased freezing behaviour as evaluated by the context-dependent and tone-dependent fear conditioning model (*P* = 0.0036, *P* = 0.0047; Fig. 2c) 24 h after training in comparison with wild-type littermates. In the short-term memory test, HDAC2KO mice showed increased freezing behaviour (*P* = 0.010; Supplementary Fig. 6e) in comparison with wild-type littermates in context-dependent conditioning. No differences in the locomotor activity or pain sensation were detected between these two groups of mice (Fig. 2d, e). Thus, loss of HDAC2 function enhanced associative learning. Furthermore, HDAC2KO mice showed a profound spatial working memory improvement in the T-maze non-matching-to-place task (*P* = 0.025, two-way ANOVA; Supplementary Fig. 6g). These data, coupled with the gain-of-function studies, suggest that HDAC2 negatively regulates memory formation in mice.

### HDAC2 regulates synaptogenesis and synaptic plasticity

Because the synapse is widely assumed to be the cellular basis for learning and memory<sup>22</sup>, we assessed whether HDAC2 regulates the density of dendritic spines and therefore synapse numbers. Indeed, the density of dendritic spines along individual dendrites of hippocampal CA1 pyramidal neurons and dentate gyrus (DG) granule cells was markedly decreased in HDAC2OE mice (*P* < 0.005; Fig. 3a and Supplementary Fig. 15a, b), but not in HDAC1OE mice (Supplementary Fig. 9a), in comparison with wild-type mice. Conversely, spine density was significantly higher in HDAC2KO mice, suggesting that HDAC2 suppresses spine formation in hippocampal CA1 and DG neurons. Detailed quantifications showed that spine shape is not related to the expression level of HDAC2 (Supplementary Fig. 15a, b). Furthermore, immunoreactivity towards synaptophysin<sup>10</sup>, which labels presynaptic terminals of functional synapses, was significantly increased in the CA1 striatum radiatum of the HDAC2KO mice (*P* < 0.0001) and decreased in HDAC2OE mice in comparison with wild-type mice (*P* = 0.0279; Fig. 3b), indicating more presynaptic terminals in



**Figure 3 | HDAC2 regulates synapse formation and plasticity in mouse hippocampus.** **a**, Representative images of Golgi staining from CA1 region of hippocampus (WT, *n* = 23; HDAC2OE, *n* = 21; HDAC2KO, *n* = 27). Yellow arrowheads indicate the spines. KO, HDAC2KO; OE, HDAC2OE. Scale bar, 10  $\mu$ m. **b**, Representative confocal images of synaptophysin (SVP)-immunoreactive (IR) signal on CA1 (HDAC2KO, *n* = 17; WT, *n* = 25; HDAC2OE, *n* = 21). py, pyramidal neuron layer; s.r., stratum radiatum. Scale bar, 50  $\mu$ m. \**P* < 0.05; \*\**P* < 0.005; \*\*\**P* < 0.001. **c**, LTP was induced by two trains of HFS (2  $\times$  100 Hz, 1 s) in the CA1 region from six-month-old HDAC2OE mice (red trace; *n* = 8) or their WT littermates (black trace; *n* = 7). By 40 min, the field excitatory postsynaptic potentials (fEPSPs) from HDAC2OE mice decayed to the baseline (103.1  $\pm$  1.8% compared with baseline), whereas fEPSPs from control mice remained potentiated (137.9  $\pm$  6.8% compared with the baseline). **d**, LTP was induced by one HFS in the CA1 region from HDAC2KO mice (red trace; *n* = 8) or their WT littermates (black trace; *n* = 6). Slices from WT mice showed a transient potentiation that decayed to the baseline (102.9  $\pm$  4.5% compared with the baseline), 40 min after induction, whereas slices from HDAC2KO mice showed a robust potentiation (144.1  $\pm$  2.0% compared with the baseline). Insets in **c** and **d** show superimposed sample sweeps from the first 5 min (black) and last 5 min (red) of the recording. Error bars show s.e.m.

HDAC2KO mice and fewer in HDAC2OE mice. The pattern of change in synaptophysin staining in the amygdala was similar to that observed in area CA1 (Supplementary Fig. 8b). Taken together, these results suggest that HDAC2 regulates synapse formation.

We next determined whether HDAC2 also regulates synaptic plasticity. To this end, we performed electrophysiological recordings on hippocampal slices prepared from HDAC2OE mice, HDAC2KO mice and their control littermates. Long-term potentiation (LTP) of CA1 neurons induced by two trains of high-frequency stimulation (HFS) on the Schaffer collaterals was observed in wild-type but impaired in HDAC2OE mice (Fig. 3c). A single train of HFS induced synaptic potentiation of WT CA1 neurons, which rapidly decayed, whereas robust LTP was readily induced by one train of HFS in CA1 neurons of HDAC2KO mice (Fig. 3d). No significant difference between wild-type and HDAC2OE groups was observed in paired-pulse facilitation (data not shown). As an additional control, we found that hippocampal LTP was not altered in HDAC1OE in comparison with wild-type controls (Supplementary Fig. 9c). Thus, overexpression of HDAC2 impaired synaptic plasticity, whereas loss of HDAC2 function facilitates synaptic plasticity. Taken together, these data indicate an inverse relationship between levels of HDAC2 and synapse numbers, which in turn affects synaptic plasticity and learning and memory.

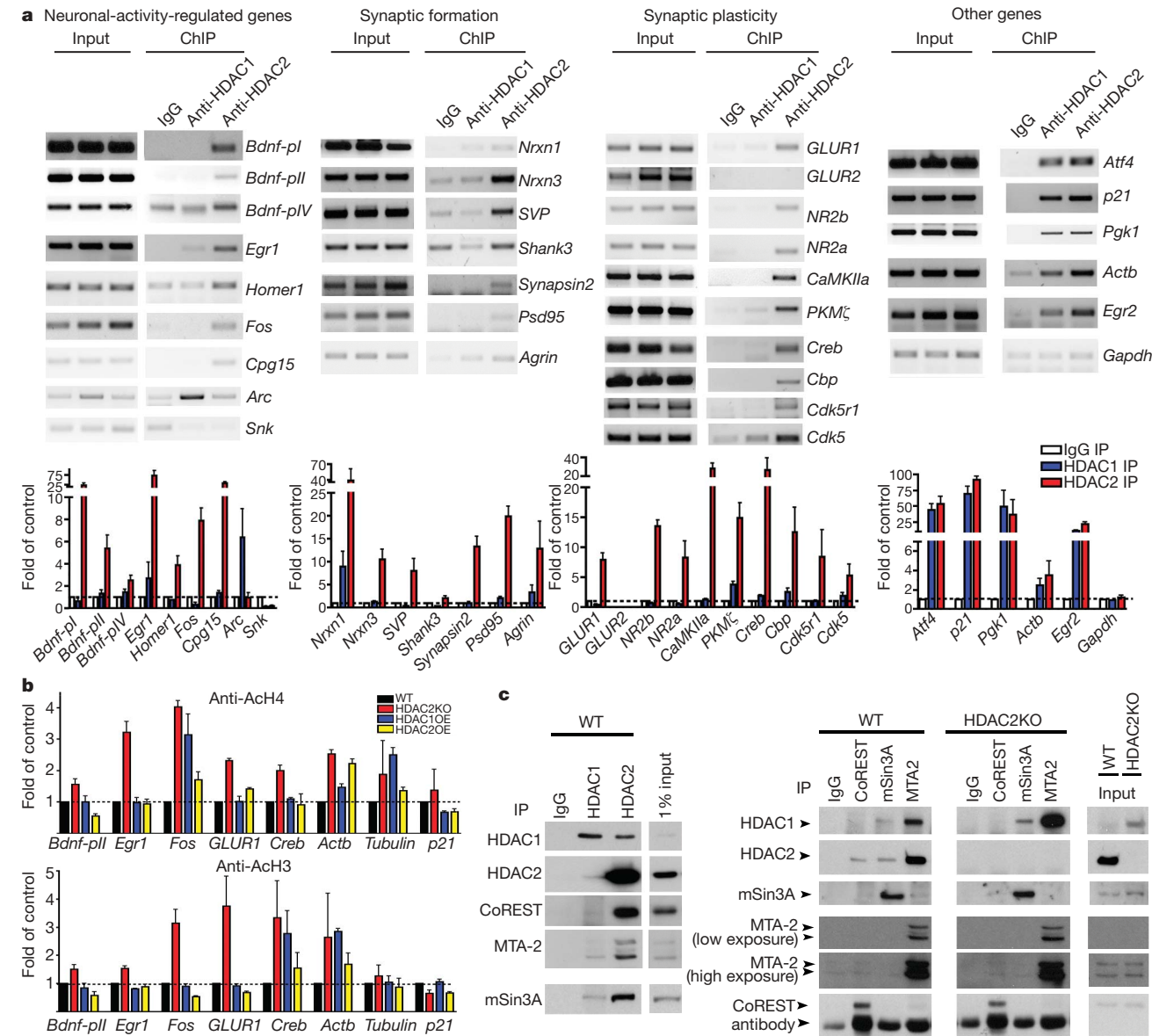
### HDAC2 suppresses neuronal gene expression

It was previously suggested that HDACs regulate learning and memory by repressing gene expression through chromatin remodelling<sup>5,7,9</sup>. Specifically, HDACs were reported to enhance memory formation



and synaptic plasticity through CREB–CBP-dependent transcriptional activation<sup>9</sup>. We speculated that HDAC2 might suppress the expression of memory-associated genes by binding to their regulatory elements. We surveyed the association of HDAC1 or HDAC2 with the promoter of a total of 24 genes implicated in synaptic remodelling/plasticity or regulated by neuronal activity (Supplementary Table 1), and other genes that are not specifically associated with synaptic function. To this end, we performed chromatin immunoprecipitation (ChIP). Consistent with previous reports that both HDAC1 and HDAC2 are associated with cell cycle genes, we observed a similar enrichment of HDAC1 and HDAC2 at the promoters of *p21* (also known as *Cdkn1a*), *Atf4* and *Pgk1* (Fig. 4a). In contrast, with minor exceptions, HDAC2 was more enriched than HDAC1 at the promoters of genes implicated in synaptic remodelling/plasticity or regulated by neuronal activity, including *Bdnf* promoter I/II, *Egr1*, *Fos*, *Cpg15* (also known as *Nrn1*), *Camk2a*, *Creb1*, *Crebbp*, *NRXN3* and the NMDA receptor subunits (Fig. 4a and Supplementary Fig. 10b). HDAC2 antibody failed to

HDAC2KO brain (Supplementary Fig. 10a), demonstrating the specificity of the previous results. Furthermore, we determined that levels of ACh3 and ACh4 were increased in multiple synaptic plasticity genes including *Bdnf-pII*, *Egr1*, *Fos* and *GLUR1* (also known as *Gria1*) promoters in the HDAC2KO mice in comparison with the wild-type controls (Fig. 4b). A significant decrease in ACh3 in *Bdnf-pII*, *Fos* and *Gria1* promoters was also detected in HDAC2OE mice, but not in HDAC1OE. Consistent with the notion that HDAC2 negatively regulates gene expression, the protein levels of multiple synaptic genes were markedly increased in the brain of HDAC2KO mice and decreased in HDAC2OE mice but not in HDAC1OE mice (Supplementary Figs 10c, d and 11a, b). Furthermore, in dissociated neurons, treatments with forskolin, KCl or BDNF (brain-derived neurotrophic factor), all of which activate the CREB pathway<sup>23</sup>, increased EGR1 expression in an EGR1–GFP reporter assay (Supplementary Figs 11c, d and 16a, b). SAHA treatment markedly potentiated the expression of EGR1–GFP, which is consistent with the notion that HDAC2 inhibition allows greater activation of the CREB–CBP pathway because



**Figure 4 | HDAC2, but not HDAC1, binds to promoters of memory-related genes.** **a**, Semi-quantitative PCR from ChIP of samples showed specificity of DNA binding for HDAC2 and HDAC1 quantified according to the real-time PCR signal ( $n = 3$ ). IP, immunoprecipitation. **b**, Changes in histone acetylation in specific gene promoter regions in brains of HDAC1OE,

HDAC2OE and HDAC2KO mice. Fragmented chromatin was immunoprecipitated with antibody recognizing ACh3 or ACh4 and quantified with real-time PCR. **c**, Nuclear extracts prepared from the WT or HDAC2KO mouse brains were immunoprecipitated to evaluate the association of HDAC1 or HDAC2 with co-repressors. Error bars show s.e.m.

EGR1 is a known CREB target gene. In these neurons, the effect of treatment with SAHA alone was not as potent as simultaneous treatment with either forskolin or KCl. Taken together, these experiments suggest that HDAC2 suppresses the expression of synaptic remodelling and plasticity genes. Thus, changes in expression of these genes in HDAC2KO and HDAC2OE mice might underlie the facilitated or impaired learning behaviour and synaptic plasticity of these mice.

### CoREST associates with HDAC2 in the brain

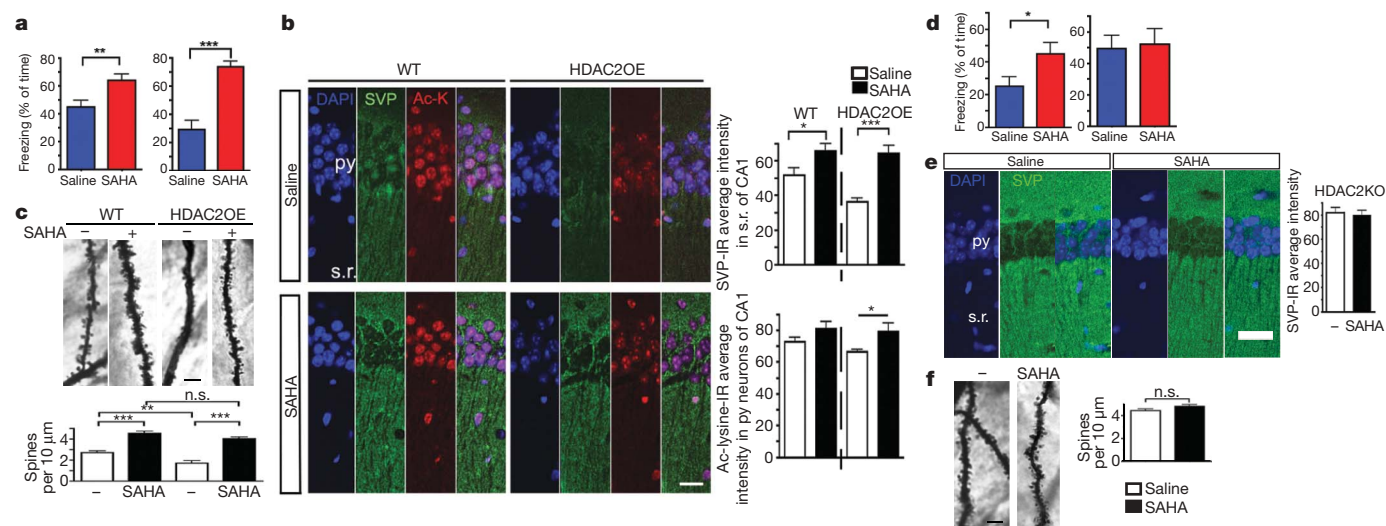
In an attempt to decipher the mechanism by which HDAC2, but not HDAC1, preferentially occupies activity-dependent gene promoters *in vivo*, we investigated the distribution of HDAC1 and HDAC2 in the brain and the association of these two HDACs with co-repressors. We found that HDAC2 was expressed more abundantly in neurons whereas HDAC1 was expressed more robustly in the cells positive for GFAP (glial fibrillary acidic protein; Supplementary Fig. 14a, b). When we evaluated the associations of HDAC1 and HDAC2 with mSin3a, CoREST and MTA2 (NuRD) in wild-type and HDAC2KO brains, we observed interactions of HDAC2 with all three co-repressors (Fig. 4c). In contrast, we failed to detect an association of HDAC1 with CoREST in wild-type mice and even in HDAC2KO brain lysates, despite the fact that HDAC1 was upregulated and its association with mSin3a and MTA2 was higher in these brains (Fig. 4c). These experiments suggest that CoREST preferentially associates with HDAC2. CoREST has been shown to have a key function in repressing neuronal gene expression by recruiting the molecular machinery responsible for silencing across a chromosomal interval<sup>24,25</sup>. Thus, our finding that HDAC2, but not HDAC1, interacts with CoREST in the brain fits in well with the notion that loss of HDAC2 function in neurons leads to the upregulation of neuronal genes involved in basal and activity-induced neuronal functions that facilitate learning and memory. Our results are also consistent with the recent report that alteration of S-nitrosylation of HDAC2 alone could affect EGR1/c-Fos expression in neurons<sup>26</sup>.

### HDACis elicit memory enhancement through HDAC2

The enhanced learning and memory, facilitated synaptic plasticity, and increased dendritic spine density and synapse number in the

HDAC2KO mice are highly reminiscent of the reported effects of treatment with HDACis<sup>5,7,10</sup>. We wondered whether HDAC2 is the major target for HDACi-induced memory enhancement. If this were so, one would expect that the learning impairment of HDAC2OE mice would be readily ameliorated by treatment with HDACis, whereas the same treatment should have no effect on memory formation in HDAC2KO mice. Because SAHA was found to be a more potent memory enhancer than sodium butyrate when given to mice (Supplementary Fig. 1c), we used SAHA in all subsequent experiments. SAHA was administered daily by intraperitoneal injection ( $25 \text{ mg kg}^{-1}$ ) for 10 days before contextual fear conditioning training and memory testing. Notably, SAHA significantly increased the freezing behaviour of HDAC2OE mice, from  $26.9 \pm 5.9\%$  to  $66.7 \pm 5.1\%$  (means  $\pm$  s.e.m. throughout; Fig. 5a). It should be noted that in the same training models, SAHA treatment increased the freezing behaviour of wild-type control mice from  $44.8 \pm 4.7\%$  to  $63.9 \pm 4.2\%$ . Thus, the freezing levels of HDAC2OE mice after SAHA treatment were comparable to those of the control mice treated with SAHA, despite the fact that saline-treated HDAC2OE mice showed lower freezing behaviour. Concordantly, SAHA treatment completely abrogated the decreased dendritic spine and synapses phenotype in HDAC2OE mice (Fig. 5b,c).

Next we investigated the effects of SAHA on HDAC2KO mice. Because HDAC2KO mice showed markedly increased freezing behaviour in comparison with wild-type littermates without treatment, we sensitized the assay by lowering the foot shock intensity from 1.0 to 0.5 mA to prevent a possible ceiling effect. In this model, SAHA treatment induced significantly higher freezing behaviour ( $P = 0.0383$ ) in the wild-type control mice ( $45.0 \pm 6.9\%$  versus  $25.0 \pm 5.8\%$ ; Fig. 5d). However, SAHA treatment did not alter the freezing behaviour in the HDAC2KO mice ( $52.1 \pm 9.8\%$  versus  $49.3 \pm 8.4\%$ ,  $P = 0.8324$ ; Fig. 5d). Furthermore, dendritic spine density of CA1 neurons and synaptophysin staining in the stratum radiatum of the HDAC2KO mice were not significantly affected by SAHA treatment (Fig. 5e, f). Consistently, although SAHA treatment modestly increased LTP in the wild-type hippocampus, it did not have a detectable effect on LTP in the HDAC2KO hippocampus (Supplementary Fig. 13). Thus,



**Figure 5 | SAHA, an HDAC inhibitor, improves associative learning by targeting HDAC2.** **a**, Memory test for mice with contextual fear-conditioning training (foot shock 1.0 mA). HDAC2OE mice (right;  $n = 12$ , each group) and WT littermates (left; SAHA group,  $n = 12$ ; saline group,  $n = 15$ ) were treated with saline or SAHA ( $25 \text{ mg kg}^{-1}$ , intraperitoneally) for 10 days before the memory test. **b**, Immunostaining images of CA1 region from WT and HDAC2OE mice that had received chronic SAHA treatment or saline treatment ( $n = 12$  for each group). Py, pyramidal neuron layer; s.r., stratum radiatum. Scale bar, 50 μm. **c**, Images of Golgi staining from the CA1 region of the hippocampus. Naive WT mice,  $n = 23$ ; SAHA-treated WT

mice,  $n = 41$ ; naive HDAC2OE mice,  $n = 21$ ; SAHA-treated mice,  $n = 32$ . Scale bar, 10 μm. **d**, Memory tests for mice with contextual fear-conditioning training (0.5-mA foot shock) after 10-day SAHA injection ( $25 \text{ mg kg}^{-1}$ , intraperitoneally). WT mice (left),  $n = 10$  in each group; HDAC2KO mice (right),  $n = 8$  in each group. **e**, Immunostaining images of CA1 region from HDAC2KO mice that had received chronic SAHA treatment or saline treatment. Saline,  $n = 15$ ; SAHA,  $n = 22$ . Scale bar, 50 μm. **f**, Images of Golgi staining of CA1 region from HDAC2KO mice. SAHA mice,  $n = 24$ ; naive mice,  $n = 27$ . \* $P < 0.05$ ; \*\* $P < 0.005$ ; \*\*\* $P < 0.001$ . Error bars show s.e.m.



HDAC2KO mice are refractory to synaptogenesis and to facilitation of synaptic plasticity and memory formation induced by SAHA. These results strongly suggest that HDAC2 is the major, if not the only, target of SAHA in eliciting memory enhancement.

## Discussion

Using mouse genetic models, we delineated the functions of HDAC1 and HDAC2 in the brain, and showed evidence that HDAC2 has a negative function in regulating memory formation. Notably, we identified HDAC2 as the major target of HDACi for facilitating learning and memory. However, we cannot rule out the possibility that other HDAC family members also have a role in synaptic function and memory, or that non-histone substrates may also be modulated by HDAC2.

Our observations suggest that HDAC1 and HDAC2 differentially regulate subsets of activity-regulated genes implicated in plasticity and memory. This is unexpected, given that HDAC1 and HDAC2 were reported to form functional heterodimers<sup>27</sup>. However, here we have shown that, HDAC2 is more abundant than HDAC1 in the co-repressor complexes containing CoREST and that HDAC2 is expressed more abundantly in neurons, providing further insight into the functional differences between HDAC1 and HDAC2 in the nervous system. Other possibilities, such as differences in post-translational modifications, might also contribute to the biochemical and functional dissociation between HDAC1 and HDAC2, because HDAC2 has recently been shown to be modified by S-nitrosylation and to regulate EGR1 expression<sup>26</sup>. It should be noted that *Hdac1* deficiency in mice is detrimental, resulting in embryonic lethality<sup>28</sup>. A recent study has indicated that loss of HDAC1 function in neurons causes DNA damage and cell death<sup>29</sup>. Conversely, *Hdac2*-deficient mice are viable and have enhanced memory formation. Using genetic models, we have found a new and unexpected function of HDAC2 as a repressor of synaptic plasticity genes and as a mediator of the beneficial effects of HDACis. Not only do these results reveal important distinct functions of HDAC1 and HDAC2 in adult mouse brain, they also suggest a new strategy for therapeutic intervention of human diseases associated with memory impairment.

## METHODS SUMMARY

The mouse HDAC1 or HDAC2 coding sequence was placed into exon 1 of the *Tau* gene. HDAC2KO mice were produced in the laboratory of R.A.D. and engineered to contain *loxP* recombination sites such that Cre-mediated recombination deleted exons 5 and 6. Sodium butyrate (Sigma) was dissolved in saline. HDAC inhibitors were dissolved in dimethylsulphoxide (DMSO) at 50 mg ml<sup>-1</sup> and diluted with saline immediately before injection (100–150 µl, intraperitoneally). Lysates for immunoblotting were prepared as described previously<sup>4</sup>. Immunoblot data were quantified by measuring the band intensity with NIH imaging software and UN-SCAN-it gel digitizing software (Silk Scientific). Immunostaining was performed as described previously<sup>4</sup> with LSMeta10 software and a confocal microscope (Zeiss). All behavioural testing was performed as described previously<sup>4</sup>. The data were analysed by unpaired Student's *t*-test. Two-way ANOVA was used to compare difference between groups at several time points.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 10 November 2008; accepted 20 February 2009.

- West, A. E. *et al.* Calcium regulation of neuronal gene expression. *Proc. Natl Acad. Sci. USA* **98**, 11024–11031 (2001).
- Guan, Z. *et al.* Integration of long-term-memory-related synaptic plasticity involves bidirectional regulation of gene expression and chromatin structure. *Cell* **111**, 483–493 (2002).
- Kurdistan, S. K. & Grunstein, M. Histone acetylation and deacetylation in yeast. *Nature Rev. Mol. Cell Biol.* **4**, 276–284 (2003).
- Goldberg, A. D., Allis, C. D. & Bernstein, E. Epigenetics: a landscape takes shape. *Cell* **128**, 635–638 (2007).
- Alarcon, J. M. *et al.* Chromatin acetylation, memory, and LTP are impaired in CBP<sup>+/-</sup> mice: a model for the cognitive deficit in Rubinstein–Taybi syndrome and its amelioration. *Neuron* **42**, 947–959 (2004).
- Korzus, E., Rosenfeld, M. G. & Mayford, M. CBP histone acetyltransferase activity is a critical component of memory consolidation. *Neuron* **42**, 961–972 (2004).

- Levenson, J. M. *et al.* Regulation of histone acetylation during memory formation in the hippocampus. *J. Biol. Chem.* **279**, 40545–40559 (2004).
- Kumar, A. *et al.* Chromatin remodeling is a key mechanism underlying cocaine-induced plasticity in striatum. *Neuron* **48**, 303–314 (2005).
- Vecsey, C. G. *et al.* Histone deacetylase inhibitors enhance memory and synaptic plasticity via CREB:CBP-dependent transcriptional activation. *J. Neurosci.* **27**, 6128–6140 (2007).
- Fischer, A. *et al.* Recovery of learning and memory is associated with chromatin remodelling. *Nature* **447**, 178–182 (2007).
- Bruserud, O. *et al.* Histone deacetylase inhibitors in cancer treatment: a review of the clinical toxicity and the modulation of gene expression in cancer cell. *Curr. Pharm. Biotechnol.* **8**, 388–400 (2007).
- Tsankova, N. M. *et al.* Sustained hippocampal chromatin regulation in a mouse model of depression and antidepressant action. *Nature Neurosci.* **9**, 519–525 (2006).
- Renthal, W. *et al.* Histone deacetylase 5 epigenetically controls behavioral adaptations to chronic emotional stimuli. *Neuron* **56**, 517–529 (2007).
- Hockley, E. *et al.* Suberoylanilide hydroxamic acid, a histone deacetylase inhibitor, ameliorates motor deficits in a mouse model of Huntington's disease. *Proc. Natl Acad. Sci. USA* **100**, 2041–2046 (2003).
- Ferrante, R. J. *et al.* Histone deacetylase inhibition by sodium butyrate chemotherapy ameliorates the neurodegenerative phenotype in Huntington's disease mice. *J. Neurosci.* **23**, 9418–9427 (2003).
- Witter, D. J. *et al.* Optimization of biaryl selective HDAC1&2 inhibitors (SHI-1:2). *Bioorg. Med. Chem. Lett.* **18**, 726–731 (2008).
- Guardiola, A. R. & Yao, T. P. Molecular cloning and characterization of a novel histone deacetylase HDAC10. *J. Biol. Chem.* **277**, 3350–3356 (2002).
- Hubbert, C. *et al.* HDAC6 is a microtubule-associated deacetylase. *Nature* **417**, 455–458 (2002).
- Salisbury, C. M. & Cravatt, B. F. Activity-based probes for proteomic profiling of histone deacetylase complexes. *Proc. Natl Acad. Sci. USA* **104**, 1171–1176 (2007).
- Roberson, E. D. *et al.* Reducing endogenous tau ameliorates amyloid  $\beta$ -induced deficits in an Alzheimer's disease mouse model. *Science* **316**, 750–754 (2007).
- Luikenhuis, S., Giacometti, E., Beard, C. F. & Jaenisch, R. Expression of MeCP2 in postmitotic neurons rescues Rett syndrome in mice. *Proc. Natl Acad. Sci. USA* **101**, 6033–6038 (2004).
- Martin, S. J., Grimwood, P. D. & Morris, R. G. Synaptic plasticity and memory: an evaluation of the hypothesis. *Annu. Rev. Neurosci.* **23**, 649–711 (2000).
- Finkbeiner, S. *et al.* CREB: a major mediator of neuronal neurotrophin responses. *Neuron* **19**, 1031–1047 (1997).
- Lunyak, V. V. *et al.* Corepressor-dependent silencing of chromosomal regions encoding neuronal genes. *Science* **298**, 1747–1752 (2002).
- Ballas, N. *et al.* REST and its corepressors mediate plasticity of neuronal gene chromatin throughout neurogenesis. *Cell* **121**, 645–657 (2005).
- Nott, A. *et al.* S-nitrosylation of histone deacetylase 2 induces chromatin remodelling in neurons. *Nature* **455**, 411–415 (2008).
- Hassig, C. A. *et al.* A role for histone deacetylase activity in HDAC1-mediated transcriptional repression. *Proc. Natl Acad. Sci. USA* **95**, 3519–3524 (1998).
- Montgomery, R. L. *et al.* Histone deacetylases 1 and 2 redundantly regulate cardiac morphogenesis, growth, and contractility. *Genes Dev.* **21**, 1790–1802 (2007).
- Kim, D. *et al.* Dereglulation of HDAC1 by p25/Cdk5 in neurotoxicity. *Neuron* **60**, 803–817 (2008).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank E. Scolnick, D. Fass, P. Sklar, T. Petryshen, B. A. Samuels, A. Fischer, C. Frank, D. Kim, S. Su and Y. Hayashi for advice and critical reading of the manuscript; T. Petryshen, A. Graybiel, J. Crittenden and M. C. Lewis for providing the T-maze behaviour model; R. Neve for providing tdTomato HSV. Funding was provided by a grant from the National Institute of Neurological Disorders and Stroke (2 RO1 NS051874) to L.-H.T., by a research fund from the Stanley Center for Psychiatric Research to L.-H.T. and S.J.H., by the National Alliance for Research on Schizophrenia and Depression Foundation to S.J.H.; by a fellowship from the Damon-Runyon Cancer Research Foundation and The Dutch Cancer Society (KWF) to J.H.D. R.J. is supported by NIH grants (5-RO1-CA087869, 5-R37-CA084198, 5-RO1-HD0445022); R.A.D. is supported by the Robert A. and Renee E. Belfer Institute for Applied Cancer Science. L.-H.T. is an investigator of the Howard Hughes Medical Institute.

**Author Contributions** L.-H.T. designed, directed and coordinated the project. J.-S.G. designed and performed the behaviour tests, biochemical assays and morphological analysis in HDACi-treated animals and genetically modified animal models. S.J.H., R.M., J.E.B. contributed to the generation and characterization of HDACis. E.G. generated HDAC1/2OE mice in R.J.'s laboratory. J.-H.D. generated HDAC2KO mice in R.A.D.'s laboratory. N.J., W.X.Y., Y.Z. and E.G. contributed to behavioural tests and biochemical analysis. J.G. performed electrophysiological analysis. T.J.F.N. performed imaging assays for cultured neurons. R.A.D., J.-H.D., E.G. and R.J. critically reviewed the experimental data. The manuscript was written by J.-S.G., S.J.H. and L.-H.T. and was commented on by all the authors.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be sent to L.-H.T. (lhtsai@mit.edu).

## METHODS

**Gene targeting construct for HDAC1OE mice.** The ~1,200-nucleotide-long mouse HDAC1 complementary DNA was amplified from a brain cDNA library and confirmed by sequencing. The cDNA was then cloned upstream of the polyadenylation (pA) signal of pC8N2 with a SpeI blunt ligation; subsequently HDAC1-pA was cloned into pBSK (Stratagene). A pGKneoLoxP sequence was directionally inserted into the XhoI–KpnI site downstream of HDAC1-pA in pBSK. HDAC1-pA-neo was released with XmaI–Acc65 and cloned in frame into exon 1 of the *Mapt* gene. The *Mapt* targeting arms were taken from pTauKR<sup>21</sup> and modified by the insertion of an XmaI and BstWI linker in the unique NcoI site. The resulting targeting vector (pTH1) containing the in-frame fusion of the HDAC1 coding sequence with exon 1 of *Mapt* was confirmed by sequencing. Mice 3–6 months old were used for the behaviour test and further analysis.

**Gene targeting construct for HDAC2OE mice.** The mouse HDAC2 cDNA was obtained by RT–PCR from mouse brain tissue. It was sequenced and subcloned into the XhoI–EcoRI site of the Topo-TA vector (Invitrogen). The pTH1 targeting vector (described above) was cut open with SmaI–SalI to release HDAC1. The HDAC2 cDNA was cut out from Topo-TA with EcoRI–XhoI and cloned into the SmaI–SalI site of pTH1, to create the pTH2 targeting vector. The in-frame fusion of HDAC2 to exon 1 of *Tau* was verified by sequencing of pTH2. The targeting vectors pTH1 and pTH2 were linearized with SacI and electroporated into the V6.5 (129 × C57BL/6) F<sub>1</sub> embryonic stem (ES) cell line. We picked 96 neomycin-resistant clones, of which 46 were analysed by Southern blots. We used only a 3' external probe, after digestion with BamHI and EcoRI. Wild-type clones show a band at 8.8 kilobases (kb). The correct targeting event results in a band shift to 13 kb for the targeted allele. Five clones were correctly targeted. Two clones were used to generate chimaeras by injections into (DBA/2 × C57BL/6) F<sub>1</sub> blastocysts. Chimaeras were mated to C57BL/6 females and offspring were analysed for germline transmission. The heterozygous knock-in strains were maintained in a mixed background and were mated to obtain homozygous animals. Mice 3–6 months old were used for the behaviour test and further analysis.

**Generation of HDAC2KO mice.** The *Hdac2* floxed allele was generated by flanking exons 5 and 6 with loxP recombination sites, ensuring the deletion of the HDAC-catalytic core of the protein after Cre-recombinase-mediated deletion (Supplementary Fig. 7a). On successful targeting of ES cells and subsequent derivation of chimaeric mice, we established a mouse strain carrying a floxed allele of *Hdac2* (*Hdac2*<sup>fl</sup>) (FVB). Infection of mouse embryonic fibroblasts (MEFs) with retroviruses expressing Cre recombinase resulted in complete ablation of HDAC2 only in MEFs carrying two floxed *Hdac2* alleles (Supplementary Fig. 7b). This indicates that the floxed *Hdac2* allele is functional and results in an *Hdac2*-null genotype on expression upon Cre recombinase. Deletion of *Hdac2* in the germline with *Elia-Cre* or *Nestin-Cre* transgenic mice resulted in viable and fertile *Hdac2*<sup>+/-</sup> mice with no obvious histological abnormalities up to one year of age. Crossing *Hdac2*<sup>+/-</sup> mice gave rise to viable *Hdac2*-deficient mice, but these mice were born with a twofold lower frequency than expected from a normal Mendelian ratio (9 *Hdac2*<sup>-/-</sup> mice out of 79 littermates, in contrast with 20 out of 79 expected; Supplementary Fig. 7c). Although *Hdac2*<sup>-/-</sup> mice are viable and are capable of producing offspring, their fertility is compromised (data not shown). *Hdac2*<sup>-/-</sup> mice (males and females) were about 25% smaller than wild-type and heterozygote littermates (data not shown). The animals used for behaviour tests were in a FVB × C57BL/6 background and mated to each other to obtain homozygous animals. Mice 3–6 months old were used for the behaviour test and further analysis. There were no difference in behaviour test results between males and females.

**Fear conditioning tests.** Context-dependent fear conditioning: training consisted of a 3-min exposure of mice to the conditioning box (context) followed by a foot shock (2 s; 0.5, 0.8 or 1.0 mA constant current). The memory test was performed 24 h later by re-exposing the mice for 3 min to the conditioning context. Freezing, defined as a lack of movement except for heartbeat and respiration associated with a crouching posture, was recorded every 10 s by two trained observers (one was unaware of the experimental conditions) for 3 min (a total of 18 sampling intervals). The number of observations indicating freezing obtained as a mean from both observers was expressed as a percentage of the total number of observations. For the short-term memory test, the memory test was performed 3 h after the foot-shock training.

**Tone-dependent fear conditioning:** training consisted of a 3-min exposure of mice to the conditioning box (context), followed by a tone (30 s, 20 kHz, 75 dB sound pressure level (SPL)) and a foot shock (2 s, 0.8 mA constant current)<sup>30</sup>. The memory test was performed 24 h later by exposing the mice for 3 min to a novel context followed by an additional 3-min exposure to a tone (10 kHz, 75 dB SPL). Freezing was recorded every 10 s by two unbiased observers as described above.

**Morris water maze test.** The water maze model<sup>31</sup> was performed in a circular tank (diameter 1.8 m) filled with opaque water. A platform (11 cm × 11 cm) was

submerged below the water's surface in the centre of the target quadrant. The swimming path of the mice was recorded by a video camera and analysed by Videomot 2 software (TSE-Systems). For each training session, the mice were placed into the maze consecutively from four random points of the tank. Mice were allowed to search for the platform for 60 s. If the mice did not find the platform within 60 s, they were gently guided to it. Mice were allowed to remain on the platform for 15 s. Two training trials were given every day; the latency for each trial was recorded for analysis. During the memory test (probe test), the platform was removed from the tank, and the mice were allowed to swim in the maze for 60 s.

**Spatial working memory on elevated T-maze.** Mice were maintained on a restricted feeding schedule at 85% of their free-feeding weight. Spatial working memory was first assessed on an elevated plastic T-maze. This consisted of a start arm (47 cm × 10 cm) and two identical goal arms (35 cm × 10 cm), surrounded by a wall 10 cm high. A plastic food well was located 3 cm from the end of each goal arm. The maze was located 1 m above the floor in a well-lit laboratory that contained various prominent distal cues beyond the maze. The mice were habituated to the maze and to drinking sweetened condensed milk, over several days before spatial non-matching-to-place testing.

Each trial consisted of a sample run and a choice run. On the sample run, the mice were forced either left or right by the presence of a plastic block, according to a pseudorandom sequence (with equal numbers of left and right turns per session, and with no more than two consecutive turns in the same direction). A reward consisting of 0.07 ml of sweetened condensed milk (diluted 1:1 with water) was available in the food well at the end of the arm. The block was then removed, and the mouse was placed, facing the experimenter, at the end of the start arm and allowed a free choice of either arm. The time interval between the sample run and the choice run was approximately 15 s. The animal was rewarded for choosing the previously unvisited arm (that is, for alternating). Mice were run one trial at a time with an intertrial interval of about 10 min. Each daily session consisted of 4 trials, and mice received 24 trials in total.

**Administration of chemicals.** SAHA was synthesized as described previously in ref. 32. WT-161 was synthesized by J.E.B. (details are available from the author). Sodium butyrate was purchased from Sigma (catalogue no. B5887). SAHA and WT-161 were dissolved in DMSO as stock solutions and diluted in saline just before injection. Sodium butyrate was prepared in saline. Mice received intraperitoneal injection daily with either SAHA or saline for either 10 or 21 days, using a vehicle consisting of 10% DMSO/90% saline (water containing 0.9% NaCl). SAHA and WT-161 solutions were first prepared by making a 50 mg ml<sup>-1</sup> compound/DMSO stock solution, then diluted 1:9 with saline (5 mg ml<sup>-1</sup> final concentration) so that the final volume of DMSO was 10%, and administered immediately with a 26-gauge needle.

**Golgi impregnation.** Golgi–Cox-stained brains<sup>33</sup> were cut to cross-sections 200 µm thick with a Vibratome (Leica) and analysed with a Zeiss 200 Axiovert microscope and Openlab software. The number of apical and basal spines on hippocampal CA1 pyramidal neurons was counted blind to the genotype. For each experimental group, a minimum of ten cells per slice (animal number *n* = 3) were analysed. CA1 hippocampal neurons within the region –1.4 mm to –1.6 mm (relative to the bregma position) were included for the analysis.

**Virus-mediated spine labelling.** Tomato-expressing herpes simplex virus (HSV; 0.5 µl; a gift from R. Neve) was stereo-injected into both sides of area CA1 or dentate gyrus at a rate of 0.05 µl min<sup>-1</sup>. Mice were killed 48 h after injection. Brains were fixed with 4% paraformaldehyde (PFA) and sectioned with a Vibratome (50-µm slices). Hippocampal slices were scanned with a confocal microscope. The image stacks were reconstructed and analysed with Image J (NIH).

**Immunohistochemistry.** Immunohistochemical analysis was performed as described previously<sup>34</sup>. Antibodies were used at a dilution of 1:1,000. Anti-HDAC1 and anti-HDAC2 antibodies were purchased from Abcam. Anti-Acetylsine, anti-Ac-H4K5, anti-Ac-H4K12, anti-Ac-H3K16, anti-CREB, anti-AKT and anti-CaMKIIα antibodies were purchased from Cell Signaling. Anti-Ac-α-tubulin (K40), anti-actin and anti-synaptophysin (SVP-38) antibodies were purchased from Sigma. Anti-NR2A and anti-NR2B were purchased from BD Biosciences. Anti-β-catenin, anti-EGR1, anti-c-Fos, anti-Brn1, anti-TLE4, anti-CDP, anti-ER81 and anti-GAPDH antibodies were purchased from Santa Cruz. Anti-NeuN antibody was purchased from Chemicon. Confocal images (1 µm) were scanned and subjected to three-dimensional reconstruction. LSMeta10 software (Zeiss) was used to calculate the mean synaptophysin intensity. Brain sections with the strongest intensity were scanned first. All other images included in the analysis were scanned with the same settings. Staining was quantified with LSMeta10 software (Zeiss).

**Protein extraction and immunoblotting.** The hippocampus and forebrain were collected and lysed in RIPA buffer. The lysates were incubated for 15 min on ice and centrifuged for 15 min at 15,000g and 4 °C. The supernatant was collected as



cytosolic protein extract. The lysates were subjected to 10% SDS–PAGE followed by immunoblotting.

**Extraction of histone proteins.** Hippocampus samples were collected and homogenized in 400 µl of TX buffer (50 mM Tris–HCl pH 8.5, 5 mM sodium butyrate). The pellets were resuspended in TX buffer containing 0.2 M HCl and incubated on ice for 30 min. Samples were centrifuged at 14,000 r.p.m. (18,000g) and the histone-containing supernatants were subjected to western analysis.

**Electrophysiological analysis.** Three-to-six-month-old HDAC2OE or HDAC2KO mice or their littermates were killed by cervical dislocation, and hippocampuses were rapidly dissected in iced oxygenated artificial cerebrospinal fluid (ACSF). Transverse hippocampal slices 400 µm thick were placed in a chamber and continuously perfused with oxygenated ACSF. A bipolar stimulating electrode (0.002-inch-diameter Nichrome wire; A-M Systems) placed in the stratum radiatum was used to elicit action potentials in CA3 axons. An ACSF-filled glass microelectrode with a resistance between 0.5 and 3 MΩ was placed in the stratum radiatum region of CA1 and was used to record the field excitatory postsynaptic potentials (fEPSPs). Data were acquired with a Heka EPC10 and analysed with PatchMaster (Heka). Peak fEPSP amplitudes from stimulators were required to be at least 2 mV, and stimulus intensity was set to produce 40% of the maximal response. Baseline responses were recorded for 20 min. fEPSPs were evoked at the CA1 synapses by stimulating Schaffer collaterals at a low frequency (2 min<sup>−1</sup>) to establish a stable baseline. Immediately after LTP induction with HFS (100 Hz, 1 s), slices from HDAC2OE and control mice showed an increase in fEPSP slope and amplitude, suggesting that short-term potentiation occurs in all groups. For HDAC2KO and its control wild-type slices, LTP was induced by applying one train of stimuli at 100 Hz for 1 s. For HDAC2OE and its control wild-type slices, LTP was induced by applying two trains of stimuli at 100 Hz for 1 s, with an interval of 20 s between trains.

**Image-based EGR1 expression assay for cultured neurons.** Embryonic cortices (E17) of EGR1–GFP BAC transgenic mice (Genesat Project) were isolated with standard procedures. Brain tissues were digested with trypsin/DNase and triturated with a glass pipe in neurobasal medium. Cortical neurons were plated at a density of 10,000 cells per well in black/clear-bottomed plates coated with poly-(D-lysine) (Costar) in neurobasal (Gibco) medium (containing 1.6% B27, 2% Glutamax, 1% penicillin/streptomycin and 5% heat-inactivated fetal calf serum) and in neurobasal medium without serum 24 h later. Under these culture conditions, the percentage of glia was estimated to be in the range 5–25%. On day 6, HDAC inhibitors or DMSO control (triplicates or quadruplicates) were added to the cultures for about 30 h. BDNF, KCl or forskolin was added to the cultures for 8 h on day 7.

Cells were fixed in 4% PFA/4% sucrose in PBS. Fixative was washed away with PBS (three wash cycles) and processed for EGR1–GFP imaging. Cells (3,000–5,000 per well) were imaged and analysed with a 5× objective using the Cellomics ArrayScan Image system. The built-in TargetActivation algorithm was optimized to measure the average EGR1–GFP expression per cell (mean fluorescence intensity per cell per well), using the Hoechst dye to mark cells. The data were normalized to control (DMSO addition).

After imaging, cells were processed for antibody staining: cells were permeabilized with 0.25% Triton X-100 (10–15 min). Triton was washed away by three wash cycles with PBS, and cells were blocked in PBS containing 10% goat serum or horse serum (1 h, 37 °C). Cells were exposed to anti-(acetyl-lysine-histone H3) or anti-(acetyl-lysine-histone H4) antibody, then washed five times with PBS followed by secondary antibody conjugated to Alexa 594, and Hoechst (1 h, room temperature (20–25 °C)). Secondary antibody was washed five times with PBS and then assayed on a Cellomics ArrayScan Image system.

**Chromatin immunoprecipitation (ChIP).** ChIP was performed with mouse forebrains fixed with 4% PFA solution and stored at −80 °C before use. Brains were chemically cross-linked by the addition of a one-tenth volume of fresh 11% formaldehyde solution for 15 min at room temperature, ten homogenized, resuspended, lysed in lysis buffers, and sonicated to solubilize and shear crosslinked DNA. Sonication conditions varied depending on the cells, culture conditions, crosslinking and equipment. We used a Misonix Sonicator 3000 and sonicated at power 7 for ten 30-s pulses (90-s pause between pulses) at 4 °C while samples were immersed in an ice bath. The resulting whole-cell extract was incubated overnight at 4 °C with 100 µl of Dynal Protein G magnetic beads that had been preincubated with 10 µg of the appropriate antibody. Beads were washed five times with RIPA buffer and once with TE (10 mM Tris–HCl pH 7.5, 1 mM EDTA) containing 50 mM NaCl. Bound complexes were eluted from the beads by being heated at 65 °C with occasional vortex-mixing, and crosslinking was reversed by incubation overnight at 65 °C. Whole-cell extract DNA (reserved from the sonication step) was also treated for crosslink reversal. Immunoprecipitated DNA and whole-cell extract DNA were then purified by treatment with RNaseA, proteinase K and multiple extractions with phenol/chloroform/3-methylbutan-1-ol. Purified DNA samples were normalized and subjected to PCR analysis. Antibodies used for pulldowns were: anti-HDAC1 (catalogue no. 31263) and anti-HDAC2

(catalogue no. 12169) from Abcam, and anti-Ach4 (catalogue no. 06-866) and anti-Ach3 (catalogue no. 06-599) from Upstate Biotechnology. After immunoprecipitation, recovered chromatin fragments were subjected to semiquantitative PCR or real-time PCR for 32–40 cycles, using primer pairs specific for 150–250-base-pair (bp) segments corresponding to mouse gene promoter regions (regions upstream of the start codon, near the first exon).

**Real-time PCR.** Real-time PCR was performed with SYBR-Green-based reagents (Express SYBR GreenER; Invitrogen), using a CFX96 real-time PCR Detection system (Bio-Rad). The relative quantities of immunoprecipitated DNA fragments were calculated by using the comparative C<sub>T</sub> method. Results were compared with a standard curve generated by serial dilutions of input DNA. Data were derived from three independent amplifications.

Primer sequences used for PCR were as follows: BDNF PI, 5'-TGATC ATCACTACGACCACG-3' and 5'-CAGCCTCTCTGAGCCAGTTACG-3'; BDNF PII, 5'-tgaggatagtgGTGGAGTTG-3' and 5'-taacttttctctctcc-3'; BDNF PIV, 5'-GCGCGGAATTCTGATTCTGGTAAT-3' and 5'-GAGAGGGCTCC ACGCTGCCTTGACG-3'; CREB, 5'-CTACACCAGCTTCCCCGGT-3' and 5'-ACGGAACAGCCGAGCTC-3'; PKMζ (100 bp upstream of the PKMζ mRNA initiation site<sup>35</sup>, which contains a cAMP response element (CRE) consensus sequence), 5'-TGTTGAGTCTGGGCCCTC-3' and 5'-CCTGGCCTCCG GACC-3'; CBP, 5'-CGGGCAGGGGATGAG-3' and 5'-GCGAGCCAGCGAG GA-3'; neurexin I, 5'-CAGGGCCTTGTCTCTGAATA-3' and 5'-GCTTTGAA TGGGGTTTTGAG-3'; neurexin III, 5'-ACTGAGAGCTAGCCACCCAGAC-3' and 5'-TTGCCCATTTGTGAATTGA-3'; PGK1, 5'-ACATTTTGGCAACCG RGAG-3' and 5'-GAAGTAGACACGTCTCACTAGTCTCGTG-3'; ATF4, 5'-GTG ATAACCTGGCAGCTTCG-3' and 5'-GGGGTAACCTGTGGCGTTAGA-3'; CaMKIIA, 5'-GACCTGGATGCTGACGAAG-3' and 5'-AGGTGATGTAGCC ATCCTG-3'; p21 (WAP/CIP1), 5'-CCACAGTTGGTCAGGGACAG-3' and 5'-CCCTCCCTCTGGGAATCTA-3'; EGR-1, 5'-GTGCCACCACTTTGG AT-3' and 5'-CGAATCGGCCTCTATTTCAA-3'; agrin, 5'-TTGTAAACCA ACAGGGGTTGC-3' and 5'-AGTTGTGGCTAGGGGAGCAC-3'; EGR-2, 5'-GGCTGCAAAATCGTTCCTG-3' and 5'-TCGGAGTATTATGGGACGT-3'; glutamate receptor 1 precursor (GluR-1/AMPA 1), 5'-GGAGGAGAGCAGA GGGAGAG-3' and 5'-TTCCTGCAATTCCTGCTTG-3'; GluR-2

5'-gcggtgctaaaatcgatgc-3' and 5'-acagagagggcgaggcagt-3'; PSD95, 5'-ccctca cctctctgagaat-3' and 5'-gaggggaaggagaaggttg-3'; homer1, 5'-ctgctgagtgtcgtg gaag-3' and 5'-atgattcactcgcgtgac-3'; p35, 5'-gagggagggcgctgagg-3' and 5'-gc agctaggagcttctgtcc-3'; cdk5, 5'-cgagcctgttgagcttgc-3' and 5'-cgcttgacagc aggtgta-3'; shank3, 5'-ttttccaggtccagtggtg-3' and 5'-ctgccccagctgtcactcc-3'; SVF, 5'-Ctagctcccgatgaatg-3' and 5'-cagcagcagcatcagaatg-3'; synapsin2, 5'-ggctttctctccctcacac-3' and 5'-tgttagcaggaggcagtg-3'; β-actin, 5'-cccatgcc aaaactctca-3' and 5'-ggccactcgagccataaaag-3'; GAPDH, 5'-ctccagggaag acctgctt-3' and 5'-ggaacaggaggagcagaga-3'; ARC, 5'-cagcataaatagcgcgtg-3' and 5'-GAGTGTGGCAGGCTCGTC-3'; Fos, 5'-gaaagcctgggctgtagat-3' and 5'-cctcagctgcgcctttat-3'; CPG15, 5'-gcgagattcgttgagatcg-3' and 5'-gggatgacac ggattgattt-3'; SNK, 5'-tttccacgtccaaagtgc-3' and 5'-gcagcgaagctttaaatacgc-3'; NR2A, 5'-tcgggttgactgatactg-3' and 5'-aggatagactgcctcgac-3'; NR2B, 5'-cct taggaagggagcgtt-3' and 5'-ggcaattaagggttggttc-3'; tubulin, 5'-TAGAACCT TCCTGCGGTCGT-3' and 5'-TTTTCTTCTGGGCTGTCTC-3'.

**Immunoprecipitation.** For preparation of nuclei, fresh mouse forebrains were washed with PBS, homogenized in 0.32 M sucrose, 1 mM MgCl<sub>2</sub>, 0.5 mM CaCl<sub>2</sub>, 1 mM NaHCO<sub>3</sub>, and then centrifuged at 710g for 10 min to obtain the nuclear pellet. The nuclear envelope was removed by the addition of 1% Triton X-100 and extracted with lysis buffer (1 mM EDTA, 0.5 mM EGTA, 10 mM Tris–HCl pH 8, 100 mM NaCl, 0.1% sodium deoxycholate, 0.5% N-lauroylsarcosine). All buffers contained complete protease inhibitor (Roche Molecular Biochemicals). For immunoprecipitation, 2–5 µg of antibody was added to the nuclear lysates and incubated for 1 h at 4 °C, then 60 µl of Protein G-Sepharose (Product Code 17-0618-01) (GE Healthcare Bio-Sciences AB) were added and incubated for 4 h. Beads were centrifuged at 2,000 r.p.m. (800g) and washed four times with 1 ml of RIPA buffer (50 mM HEPES pH 7.6, 1 mM EDTA, 0.7% sodium deoxycholate, 1% Nonidet P40, 0.5 M LiCl). Proteins were eluted with sample buffer (containing 1% SDS and 2% 2-mercaptoethanol) and analysed.

Antibodies used were HDAC1, HDAC2 and mSin3A from Abcam, and MTA-2 and CoREST from Upstate Biotechnology.

- Radulovic, J., Ruhmann, A., Liepold, T. & Spiess, J. Modulation of learning and anxiety by corticotropin-releasing factor (CRF) and stress: differential roles of CRF receptors 1 and 2. *J. Neurosci.* **19**, 5016–5025 (1999).
- Morris, R. G., Garrud, P., Rawlins, J. N. & O'Keefe, J. Place navigation impaired in rats with hippocampal lesions. *Nature* **297**, 681–683 (1982).
- Breslow, R., Marks, P. & Rifkind, R. and Jursic, B. *Novel Potent Inducers of Terminal Differentiation and Methods Thereof*. US Patent WO/1993/007148 (1993).
- Ramon-Molinier, E. *Contemporary Research Methods in Neuroanatomy Synaptic Plasticity: Multiple Forms and Mechanisms* (Springer, 1970).

34. Guan, J. S. *et al.* Interaction with vesicle luminal protachykinin regulates surface expression of  $\delta$ -opioid receptors and opioid analgesia. *Cell* **122**, 619–631 (2005).
35. Hernandez, A. I. *et al.* Protein kinase M $\zeta$  synthesis from a brain mRNA encoding an independent protein kinase C $\zeta$  catalytic domain. Implications for the molecular mechanism of memory. *J. Biol. Chem.* **278**, 40305–40316 (2003).



# Spatial correlation between submillimetre and Lyman- $\alpha$ galaxies in the SSA 22 protocluster

Yoichi Tamura<sup>1,2</sup>, Kotaro Kohno<sup>3</sup>, Kouichiro Nakanishi<sup>2,4</sup>, Bunyo Hatsukade<sup>3</sup>, Daisuke Iono<sup>3,4</sup>, Grant W. Wilson<sup>5</sup>, Min S. Yun<sup>5</sup>, Tadafumi Takata<sup>2</sup>, Yuichi Matsuda<sup>2</sup>, Tomoka Tosaki<sup>4</sup>, Hajime Ezawa<sup>4</sup>, Thushara A. Perera<sup>5</sup>, Kimberly S. Scott<sup>5</sup>, Jason E. Austermann<sup>5</sup>, David H. Hughes<sup>6</sup>, Itziar Aretxaga<sup>6</sup>, Aeree Chung<sup>5</sup>, Tai Oshima<sup>4</sup>, Nobuyuki Yamaguchi<sup>4</sup>, Kunihiro Tanaka<sup>4</sup> & Ryohei Kawabe<sup>4</sup>

Lyman- $\alpha$  emitters are thought to be young, low-mass galaxies with ages of  $\sim 10^8$  yr (refs 1, 2). An overdensity of them in one region of the sky (the SSA 22 field) traces out a filamentary structure in the early Universe at a redshift of  $z \approx 3.1$  (equivalent to 15 per cent of the age of the Universe) and is believed to mark a forming protocluster<sup>3,4</sup>. Galaxies that are bright at (sub)millimetre wavelengths are undergoing violent episodes of star formation<sup>5–8</sup>, and there is evidence that they are preferentially associated with high-redshift radio galaxies<sup>9</sup>, so the question of whether they are also associated with the most significant large-scale structure growing at high redshift (as outlined by Lyman- $\alpha$  emitters) naturally arises. Here we report an imaging survey of 1,100- $\mu\text{m}$  emission in the SSA 22 region. We find an enhancement of submillimetre galaxies near the core of the protocluster, and a large-scale correlation between the submillimetre galaxies and the low-mass Lyman- $\alpha$  emitters, suggesting synchronous formation of the two very different types of star-forming galaxy within the same structure at high redshift. These results are in general agreement with our understanding of the formation of cosmic structure.

Many different populations of young star-forming galaxies in the early Universe are known, but the relations among them and to the cosmic large-scale structure are still not well understood. The members of one of these populations are characterized by their strong Lyman- $\alpha$  ( $\text{Ly}\alpha$ ) emission (luminosity,  $L_{\text{Ly}\alpha} \gtrsim 10^{42} \text{ erg s}^{-1}$ ), arising from ionized gas; their deficiency in ultraviolet continuum emission, which is interpreted as having a relatively small stellar component<sup>1</sup> ( $M_{\text{star}} \lesssim 10^9 M_{\odot}$ , where  $M_{\odot}$  is the solar mass); and their small size<sup>2</sup> ( $\lesssim 1 \text{ kpc}$  in diameter). The  $\text{Ly}\alpha$  emitters towards SSA 22 trace a large-scale ( $\sim 10 \text{ arcmin}$ ) filamentary structure that extends over several tens of megaparsecs (co-moving scale) and which may be the largest protocluster yet detected at high redshift<sup>4</sup>.

Massive galaxies forming through accretion and mergers of small galaxies in such high-density environments are expected to be dust-obscured starbursts, which are too faint to detect at optical wavelengths but are observed as submillimetre-bright galaxies (SMGs). It is known from previous studies that SMGs have molecular gas reservoirs of  $10^{10} M_{\odot}$ – $10^{11} M_{\odot}$  (ref. 10) for their star-formation activities, suggesting that they are progenitors of massive elliptical galaxies seen in the cores of present-day clusters<sup>8,11</sup>. Individual  $\sim 5\text{-arcmin}^2$ -wide, deep submillimetre surveys in the direction of powerful, high-redshift radio galaxies, which are also believed to trace protoclusters<sup>12</sup>, have presented tentative evidence for an enhancement in the number density of submillimetre sources around them<sup>9</sup>. Although these observations were

limited in sensitivity and spatial coverage, they support the idea that SMGs are related to large-scale structure. To better understand the connection between the formation of massive galaxies and large-scale structure, we mapped the large-scale distribution of (sub)millimetre-bright, dusty starburst galaxies in the SSA 22 protocluster.

We carried out a wide-area ( $390\text{-arcmin}^2$ ) survey of the SSA 22 field at 1,100  $\mu\text{m}$  using the AzTEC camera<sup>13</sup> mounted on the Atacama Submillimeter Telescope Experiment (ASTE)<sup>14</sup>, Chile (see also Supplementary Fig. 1). Our AzTEC map (Fig. 1a), which is more than 20 times larger than any of the existing maps at submillimetre wavelengths in this field (see, for example, refs 15–17), is wide enough to cover the region of the entire protocluster. We have detected 30 SMGs with signal-to-noise ratios  $s/n \geq 3.5$  (a full source list is given in Supplementary Table 1). Their intrinsic flux densities are in the range 1.9–8.4 mJy ( $1 \text{ Jy} = 10^{-23} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ Hz}^{-1}$ ), corresponding to far-infrared luminosities of  $L_{\text{FIR}} > 4 \times 10^{12} L_{\odot}$  (where  $L_{\odot}$  is the solar luminosity) if we assume an emissivity index of  $\beta = 1.5$ , a dust temperature of  $T_{\text{dust}} = 40 \text{ K}$  and that the sources are located at  $z = 2$ – $6$ . The inferred star-formation rates of the 1,100- $\mu\text{m}$  sources are  $\sim 10^3 M_{\odot} \text{ yr}^{-1}$ , assuming that star formation is the dominant mechanism that heats the dust.

The most prominent new finding is that the distribution of the brighter ( $\geq 2.7 \text{ mJy}$ ) half of the 1,100- $\mu\text{m}$  sources (15 of the 30, hereafter termed ‘bright’ SMGs; Table 1), which suffer little from incompleteness and false detections (Supplementary Figs 2 and 3), appears to be correlated with the high-density region of  $\text{Ly}\alpha$  emitters<sup>4</sup>, as seen in Fig. 1b. A concentration of bright SMGs  $\sim 5 \text{ arcmin}$  northwest of the field centre is evident. Seven of the 15 bright SMGs (47%) are concentrated within a  $50\text{-arcmin}^2$  region in the direction that has a large-scale filamentary structure of  $\text{Ly}\alpha$  emitters  $\sim 50 \text{ Mpc}$  in depth (see Fig. 1 of ref. 18). The number density over this region is 2–3 times higher than those found in blank-field surveys at 1,100  $\mu\text{m}$  (ref. 19). Furthermore, the three most significant sources ( $8.4^{+0.8}_{-1.0}$ ,  $4.4^{+0.9}_{-0.8}$ ,  $4.1^{+1.0}_{-0.8} \text{ mJy}$ ) are all located close ( $< 4.5 \text{ arcmin}$ ) to the peak of the  $\text{Ly}\alpha$  emitter overdensity. Photometric redshift estimates for the SMGs based on their radio and 24–1,100- $\mu\text{m}$  flux ratios (Supplementary Fig. 4) indicate that they are probably at high redshift ( $z > 1$ ). The redshift estimates also suggest that some fraction of the bright SMGs, including the three most significant sources towards SSA 22, can be located at  $z = 3.1$  and may mark the local peak of the underlying mass distribution in the protocluster.

A two-point angular cross-correlation function is often used in determining the fractional increase in the probability of finding a

<sup>1</sup>Department of Astronomy, University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. <sup>2</sup>National Astronomical Observatory of Japan, 2-21-1 Osawa, Mitaka, Tokyo 181-8588, Japan. <sup>3</sup>Institute of Astronomy, University of Tokyo, 2-21-1 Osawa, Mitaka, Tokyo 181-0015, Japan. <sup>4</sup>Nobeyama Radio Observatory, National Astronomical Observatory of Japan, Minamimaki, Minamisaku, Nagano 384-1305, Japan. <sup>5</sup>Department of Astronomy, University of Massachusetts, Amherst, Massachusetts 01003, USA. <sup>6</sup>Instituto Nacional de Astrofísica, Óptica y Electrónica, Aptdo. Postal 51 y 216, 72000 Puebla, Puebla, Mexico.

source of one population within a unit solid angle as a function of angular distance from a source of another population, relative to a random distribution. We use an angular cross-correlation function to quantify the degree of cohabitation between the Ly $\alpha$  emitters and the bright SMGs. Figure 2 shows the cross-correlation function, which reveals strong correlation signals at angular distances less than 5 arcmin for the bright sample, suggesting close association of the Ly $\alpha$  emitters with the bright SMGs that are most probably embedded in the more massive dark haloes. Monte Carlo simulations (Supplementary Information) also show a correspondence between the two distributions, at a 97.3% significance level, further supporting the positional association of Ly $\alpha$  emitters with bright SMGs.

The gravitational lensing magnification of background galaxies by foreground large-scale structure would immediately preclude the

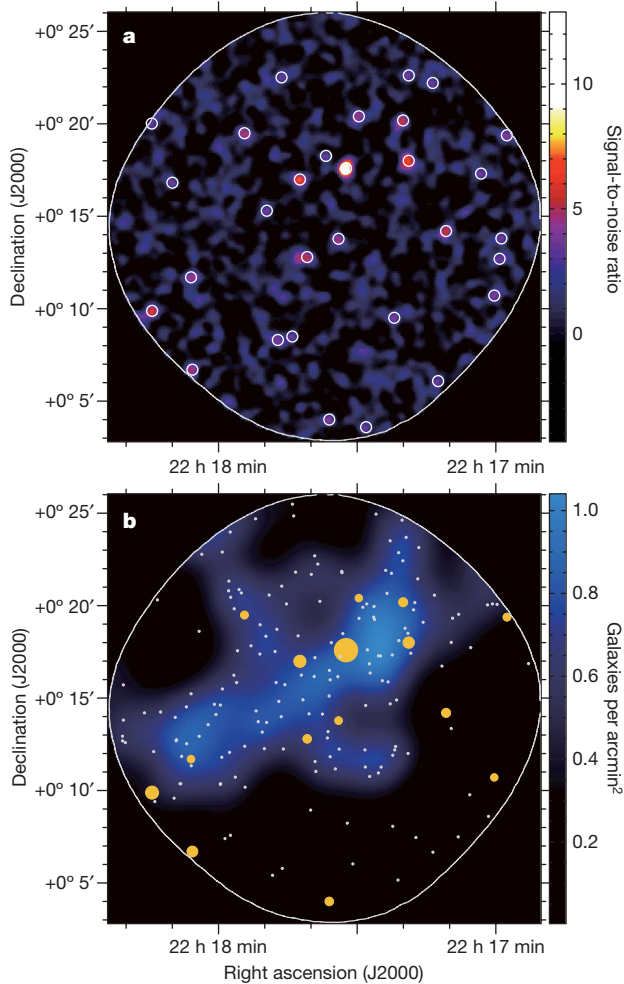
**Table 1 | The bright SMG sample found in SSA 22**

| Source name   | Coordinate (J2000) |                | Flux density (mJy) |                                     | s/n  |
|---------------|--------------------|----------------|--------------------|-------------------------------------|------|
|               | RA (h:min:s)       | Dec.           | Observed*          | S <sub>deboost</sub> †              |      |
| SSA22-AzTEC1  | 22:17:32.42        | +0° 17' 35.5'' | 8.7 ± 0.7          | 8.4 <sup>+0.8</sup> <sub>-1.0</sub> | 12.8 |
| SSA22-AzTEC2  | 22:17:42.38        | +0° 16' 59.3'' | 4.9 ± 0.7          | 4.4 <sup>+0.9</sup> <sub>-0.8</sub> | 7.2  |
| SSA22-AzTEC3  | 22:17:18.85        | +0° 18' 0.0''  | 4.7 ± 0.7          | 4.1 <sup>+1.0</sup> <sub>-0.8</sub> | 6.8  |
| SSA22-AzTEC4  | 22:18:14.37        | +0° 9' 53.1''  | 5.6 ± 0.9          | 4.7 <sup>+1.2</sup> <sub>-1.0</sub> | 6.2  |
| SSA22-AzTEC5  | 22:17:10.77        | +0° 14' 11.8'' | 4.0 ± 0.7          | 3.3 <sup>+1.0</sup> <sub>-0.8</sub> | 5.6  |
| SSA22-AzTEC6  | 22:17:20.07        | +0° 20' 11.0'' | 4.0 ± 0.7          | 3.3 <sup>+1.0</sup> <sub>-0.8</sub> | 5.6  |
| SSA22-AzTEC7  | 22:17:40.82        | +0° 12' 47.6'' | 3.6 ± 0.7          | 3.1 <sup>+0.8</sup> <sub>-0.9</sub> | 5.2  |
| SSA22-AzTEC8  | 22:18:5.65         | +0° 6' 42.0''  | 4.9 ± 1.0          | 3.9 <sup>+1.2</sup> <sub>-1.2</sub> | 5.0  |
| SSA22-AzTEC9  | 22:17:54.40        | +0° 19' 29.5'' | 3.6 ± 0.7          | 2.9 <sup>+0.9</sup> <sub>-0.9</sub> | 5.0  |
| SSA22-AzTEC10 | 22:17:34.03        | +0° 13' 46.8'' | 3.4 ± 0.7          | 2.8 <sup>+0.9</sup> <sub>-0.9</sub> | 4.8  |
| SSA22-AzTEC11 | 22:17:29.64        | +0° 20' 24.4'' | 3.3 ± 0.7          | 2.7 <sup>+0.9</sup> <sub>-0.9</sub> | 4.7  |
| SSA22-AzTEC12 | 22:17:36.04        | +0° 4' 0.2''   | 4.0 ± 0.9          | 3.1 <sup>+1.1</sup> <sub>-1.1</sub> | 4.5  |
| SSA22-AzTEC13 | 22:18:5.95         | +0° 11' 41.9'' | 3.3 ± 0.7          | 2.7 <sup>+0.9</sup> <sub>-1.0</sub> | 4.5  |
| SSA22-AzTEC14 | 22:17:0.34         | +0° 10' 42.6'' | 3.7 ± 0.9          | 2.7 <sup>+1.2</sup> <sub>-1.2</sub> | 4.2  |
| SSA22-AzTEC15 | 22:16:57.60        | +0° 19' 22.8'' | 4.1 ± 1.0          | 2.9 <sup>+1.4</sup> <sub>-1.3</sub> | 4.2  |

A full list of the 30 submillimetre galaxies is given in Supplementary Table 1. The astrometric accuracy of the catalogue is ~10 arcsec.

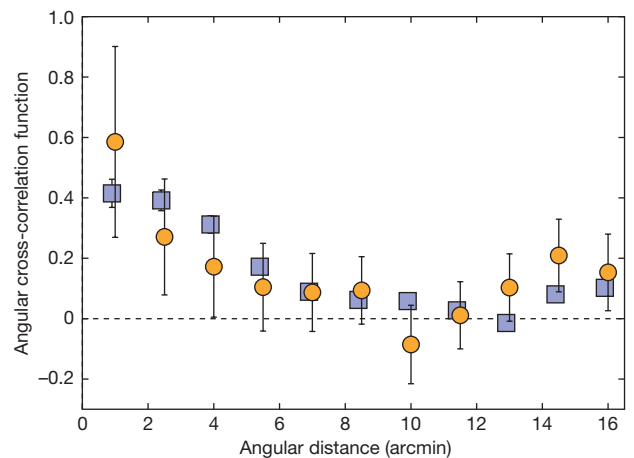
\* Observed flux density before flux bias correction, plus the 1 $\sigma$  error.

† Deboosted flux density (flux density corrected for the flux bias due to confusion noise using the method described elsewhere<sup>28</sup>), plus the 68% confidence interval.



**Figure 1 | The positions of 1,100- $\mu$ m sources and Ly $\alpha$  emitters towards the SSA 22 protocluster region.** **a**, The colour scale shows the map of signal-to-noise ratio at 1,100  $\mu$ m. The map shows 30 sources with signal-to-noise ratios  $\geq 3.5$  (circles). Observations of SSA 22 (field centre at RA = 22 h 17 min 36 s, dec. = +0° 15' 00'' (J2000)) were obtained using the AzTEC camera<sup>13</sup>, operating at 1,100  $\mu$ m, mounted on the ASTE 10-m submillimetre telescope<sup>14</sup> during the July–September 2007 observing season. The data consist of a total of 42 h of integration time on source under excellent conditions (zenith atmospheric opacity at 220 GHz,  $\tau_{220 \text{ GHz}} = 0.01\text{--}0.10$ ). This resulted in a root-mean-square noise level of 0.68–0.99 mJy per beam over 390 arcmin<sup>2</sup>. The point spread function of AzTEC on ASTE has a full-width at half-maximum of  $28 \pm 1$  arcsec. **b**, The locations of the bright submillimetre galaxies with  $S_{1,100 \mu\text{m}} \geq 2.7$  mJy (orange filled circles) and the Ly $\alpha$  emitters at  $z = 3.1$  (white dots). The sizes of the orange circles are proportional to their 1,100  $\mu$ m fluxes. The number density field of the Ly $\alpha$  emitters is shown in the colour scale, highlighting the density enhancement of the Ly $\alpha$  emitters, which is thought to trace out the underlying large-scale structure at  $z = 3.1$ .

physical connection between the galaxies and the foreground structure. Some authors<sup>20,21</sup> have reported correlations between bright (sub)millimetre sources and optically selected low-redshift galaxies (mostly at  $z < 1$ ) in other regions of the sky. In general, SMGs are often found at high redshift (median,  $z = 2.2$ ; ref. 22), and the maximal gravitational lensing magnification for a background galaxy at  $z \gtrsim 2$  occurs when the foreground lensing structure is at  $z \approx 0.5$ . Therefore, they concluded that the correlation signal is most probably the result of amplification of background SMGs due to gravitational weak lensing by the foreground low-redshift galaxies. By contrast, the origin of the correlation signals in SSA 22 is most likely intrinsic to the large-scale structure in which both populations, SMGs and Ly $\alpha$  emitters, are embedded. Because the redshift estimates for the SMGs place them at distances coeval with the Ly $\alpha$  emitters, it is unlikely that the correlation seen in SSA 22 is due to amplification of a much higher-redshift ( $z \gg 3.1$ ) SMG population lensed by the structure traced by the Ly $\alpha$  emitters, which are all located at  $z = 3.1$  (not  $z \approx 0.5$ ).



**Figure 2 | Angular cross-correlation between submillimetre galaxies and Ly $\alpha$  emitters.** The two-point angular cross-correlation function shown here is computed for the 166 Ly $\alpha$  emitters and the 15 brightest ( $S_{1,100 \mu\text{m}} \geq 2.7$  mJy) submillimetre galaxies (orange circles). For reference, we also show the two-point angular autocorrelation function for the SSA 22 Ly $\alpha$  emitters (blue squares). Small-number statistics prevent us from constraining the auto-correlation function well for the submillimetre galaxies. The correlation functions are computed using the estimator of ref. 29. The error bars are estimated from the root mean square of 1,000 bootstrap samples. See Supplementary Information for details.



We do not detect the dust emission from individual Ly $\alpha$  emitters at the sensitivity of our 1,100- $\mu$ m observations. This is a strong indication that SMGs and Ly $\alpha$  emitters are different populations, even though the Ly $\alpha$  emitters are spatially correlated with the SMGs. Of the 166 Ly $\alpha$  emitters within our 1,100- $\mu$ m coverage, none are within the  $2\sigma$  error circle ( $\sim 26$ -arcsec diameter for  $3.5 < s/n < 4.5$  and  $\lesssim 20$  arcsec for  $s/n > 4.5$ ) of an SMG; on average, we expect 2–3 SMGs to have a chance to be associated with a Ly $\alpha$  emitter in AzTEC's 28-arcsec beam if 30 SMGs and 166 Ly $\alpha$  emitters are randomly scattered in the 390-arcmin<sup>2</sup> region of our survey. To estimate the dust mass of a typical Ly $\alpha$  emitter in SSA 22, we stack the 1,100- $\mu$ m images on the positions of the 166 Ly $\alpha$  emitters. We see no dust emission above 107  $\mu$ Jy ( $2\sigma$ ) at 1,100  $\mu$ m, and derive limits on far-infrared luminosity of  $L_{\text{FIR}} < 1.9 \times 10^{11} L_{\odot}$  and  $L_{\text{FIR}} < 1.7 \times 10^{12} L_{\odot}$  for  $\beta = 1.5$  and, respectively,  $T_{\text{dust}} = 40$  K and  $T_{\text{dust}} = 70$  K. These luminosities correspond to respective dust masses of  $M_{\text{dust}} < 1.4 \times 10^7 M_{\odot}$  and  $M_{\text{dust}} < 5.8 \times 10^6 M_{\odot}$ , assuming a dust emissivity of  $\kappa_{850 \mu\text{m}} = 0.15 \text{ m}^2 \text{ kg}^{-1}$  (ref. 23). This limit is 3–40 times lower than the dust masses previously derived<sup>24,25</sup> for Ly $\alpha$  emitters at  $z = 6.5$ . Of course, the result from a simple stacking analysis cannot strongly constrain the dust properties of the Ly $\alpha$  emitter population. Nevertheless, this limit is 1–2 orders of magnitude lower than the average dust mass found in the population of SMGs, supporting the argument that Ly $\alpha$  emitters are on average less dust obscured<sup>1</sup> than SMGs.

These results provide evidence in favour of the synchronous formation of two very different types of high-redshift star-forming galaxy, SMGs and Ly $\alpha$  emitters, within the same cosmic structure. Although the formation process of SMGs is not yet fully understood, the observational evidence shown here suggests that they may form preferentially in regions of high mass concentration, which is consistent with predictions from the standard model of hierarchical structure formation<sup>26,27</sup>: we are presumably observing a galaxy-formation site where large-scale accumulation of baryonic matter is occurring within the large dark matter halo. Millimetre/submillimetre interferometric identifications followed by accurate measurements of the SMG redshifts will allow us to investigate this further.

Received 23 August 2008; accepted 3 March 2009.

- Gawiser, E. *et al.* Ly $\alpha$ -emitting galaxies at  $z = 3.1$ :  $L^*$  progenitors experiencing rapid star formation. *Astrophys. J.* **671**, 278–284 (2007).
- Pirzkal, N. *et al.* Optical-to-mid-infrared observations of Ly $\alpha$  galaxies at  $z \approx 5$  in the Hubble Ultra Deep Field: a young and low-mass population. *Astrophys. J.* **667**, 49–59 (2007).
- Steidel, C. C. *et al.* A large structure of galaxies at redshift  $z \sim 3$  and its cosmological implications. *Astrophys. J.* **492**, 428–438 (1998).
- Hayashino, T. *et al.* Large-scale structure of emission-line galaxies at  $z = 3.1$ . *Astron. J.* **128**, 2073–2079 (2004).
- Smail, I., Ivison, R. J. & Blain, A. W. A deep sub-millimetre survey of lensing clusters: a new window on galaxy formation and evolution. *Astrophys. J.* **490**, L5–L8 (1997).
- Hughes, D. H. *et al.* High-redshift star formation in the Hubble Deep Field revealed by a submillimetre-wavelength survey. *Nature* **394**, 241–247 (1998).
- Barger, A. J. *et al.* Submillimetre-wavelength detection of dusty star-forming galaxies at high redshift. *Nature* **394**, 248–251 (1998).
- Blain, A. W. *et al.* Submillimetre galaxies. *Phys. Rep.* **369**, 111–176 (2002).
- Stevens, J. A. *et al.* The formation of cluster elliptical galaxies as revealed by extensive star formation. *Nature* **425**, 264–267 (2003).
- Solomon, P. M. & Vanden Bout, P. A. Molecular gas at high redshift. *Annu. Rev. Astron. Astrophys.* **43**, 677–725 (2005).

- Eales, S. *et al.* The Canada-UK deep submillimetre survey: first submillimetre images, the source counts, and resolution of the background. *Astrophys. J.* **515**, 518–524 (1999).
- Le Fèvre, O. *et al.* Clustering around the radio galaxy MRC 0316–257 at  $z = 3.14$ . *Astrophys. J.* **471**, L11–L14 (1996).
- Wilson, G. W. *et al.* The AzTEC mm-wavelength camera. *Mon. Not. R. Astron. Soc.* **386**, 807–818 (2008).
- Ezawa, H., Kawabe, R., Kohno, K. & Yamamoto, S. The Atacama Submillimetre Telescope Experiment (ASTE). *Proc. SPIE* **5489**, 763–772 (2004).
- Chapman, S. C. *et al.* The properties of microjansky radio sources in the Hubble Deep Field-North, SSA 13, and SSA 22 Fields. *Astrophys. J.* **585**, 57–66 (2003).
- Blain, A. W. *et al.* Clustering of submillimetre-selected galaxies. *Astrophys. J.* **611**, 725–731 (2004).
- Scott, S. *et al.* A combined re-analysis of existing blank-field SCUBA surveys: comparative 850- $\mu$ m source lists, combined number counts, and evidence for strong clustering of the bright submillimetre galaxy population on arcminute scales. *Mon. Not. R. Astron. Soc.* **370**, 1057–1105 (2006).
- Matsuda, Y. *et al.* Large-scale filamentary structure around the protocluster at redshift  $z = 3.1$ . *Astrophys. J.* **634**, L125–L128 (2005).
- Laurent, G. T. *et al.* The Bolocam Lockman Hole millimetre-wave galaxy survey: galaxy candidates and number counts. *Astrophys. J.* **623**, 742–762 (2005).
- Almaini, O. *et al.* Correlations between bright submillimetre sources and low-redshift galaxies. *Mon. Not. R. Astron. Soc.* **358**, 875–882 (2005).
- Austermann, J. E. *et al.* AzTEC millimetre survey of the COSMOS field - II. Source count overdensity and correlations with large-scale structure. *Mon. Not. R. Astron. Soc.* (in the press); preprint at (<http://arxiv.org/abs/0812.0814>) (2008).
- Chapman, S. C. *et al.* A redshift survey of the submillimetre galaxy population. *Astrophys. J.* **622**, 772–796 (2005).
- Dunne, L. *et al.* A census of metals at high and low redshift and the connection between submillimetre sources and spheroid formation. *Mon. Not. R. Astron. Soc.* **341**, 589–598 (2003).
- Webb, T. M. A. *et al.* Deep submillimetre observations of two Ly $\alpha$ -emitting galaxies at  $z \sim 6.5$ . *Astrophys. J.* **659**, 76–83 (2007).
- Boone, F. *et al.* Millimetre observations of HCM 6A, a gravitationally lensed Ly $\alpha$  emitting galaxy at  $z = 6.56$ . *Astron. Astrophys.* **475**, 513–517 (2007).
- Mo, H. J. & White, S. D. M. An analytic model for the spatial clustering of dark matter haloes. *Mon. Not. R. Astron. Soc.* **282**, 347–361 (1996).
- Cole, S. *et al.* Hierarchical galaxy formation. *Mon. Not. R. Astron. Soc.* **319**, 168–204 (2000).
- Scott, K. S. *et al.* AzTEC millimetre survey of the COSMOS field - I. Data reduction and source catalogue. *Mon. Not. R. Astron. Soc.* **385**, 2225–2238 (2008).
- Landy, S. D. & Szalay, A. S. Bias and variance of angular correlation functions. *Astrophys. J.* **412**, 64–71 (1993).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We acknowledge T. Yamada and T. Hayashino for providing the Ly $\alpha$  emitter catalogue. We are grateful to H. Hirashita, T. Suwa, T. Kodama, M. Sameshima, M. Hayashi, T. T. Takeuchi and S. Komugi for discussions. We thank M. Uehara and the ASTE and AzTEC staff for their support. The ASTE project is led by Nobeyama Radio Observatory, in collaboration with the University of Chile, the University of Tokyo, Nagoya University, Osaka Prefecture University, Ibaraki University, and Hokkaido University. This work is based in part on archival data obtained with the NASA Spitzer Space Telescope.

**Author Contributions** K.N., Y.T., T. Takata, K.K. and R.K. designed and proposed the survey. Y.T., K.K., K.N., B.H., D.I. and T. Tosaki conducted the observing runs for two months. G.W.W., T.A.P., J.E.A. and K.S.S. developed the AzTEC instrument and the fundamental AzTEC reduction pipeline. H.E., D.H.H., I.A., T.O., N.Y. and K.T. contributed to the operation of AzTEC and ASTE during the survey. Y.T. and B.H. processed the raw AzTEC data, carried out simulations to create a source catalogue and computed the correlation functions. M.S.Y. and A.C. processed the Very Large Array 20-cm data. Y.M. provided the Ly $\alpha$  emitter catalogue and contributed to discussions, especially on Ly $\alpha$  emitters. All the authors discussed the results.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to Y.T. ([yoichi.tamura@nao.ac.jp](mailto:yoichi.tamura@nao.ac.jp)).

## LETTERS

# A large iron isotope effect in $\text{SmFeAsO}_{1-x}\text{F}_x$ and $\text{Ba}_{1-x}\text{K}_x\text{Fe}_2\text{As}_2$

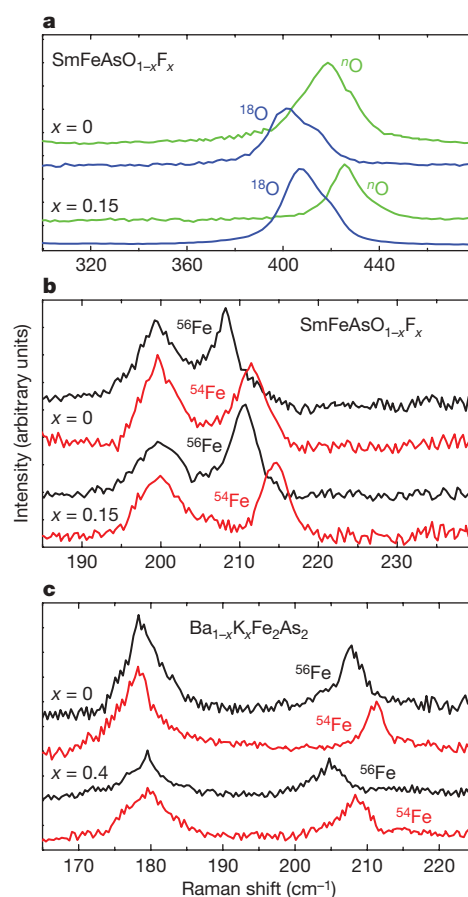
R. H. Liu<sup>1</sup>, T. Wu<sup>1</sup>, G. Wu<sup>1</sup>, H. Chen<sup>1</sup>, X. F. Wang<sup>1</sup>, Y. L. Xie<sup>1</sup>, J. J. Ying<sup>1</sup>, Y. J. Yan<sup>1</sup>, Q. J. Li<sup>1</sup>, B. C. Shi<sup>1</sup>, W. S. Chu<sup>2,3</sup>, Z. Y. Wu<sup>2,3</sup> & X. H. Chen<sup>1</sup>

The recent discovery of superconductivity in oxypnictides with a critical transition temperature ( $T_C$ ) higher than the McMillan limit of 39 K (the theoretical maximum predicted by Bardeen–Cooper–Schrieffer theory) has generated great excitement<sup>1–5</sup>. Theoretical calculations indicate that the electron–phonon interaction is not strong enough to give rise to such high transition temperatures<sup>6</sup>, but strong ferromagnetic/antiferromagnetic fluctuations have been proposed to be responsible<sup>7–9</sup>. Superconductivity and magnetism in pnictide superconductors, however, show a strong sensitivity to the crystal lattice, suggesting the possibility of unconventional electron–phonon coupling. Here we report the effect of oxygen and iron isotope substitution on  $T_C$  and the spin-density wave (SDW) transition temperature ( $T_{\text{SDW}}$ ) in the  $\text{SmFeAsO}_{1-x}\text{F}_x$  and  $\text{Ba}_{1-x}\text{K}_x\text{Fe}_2\text{As}_2$  systems. The oxygen isotope effect on  $T_C$  and  $T_{\text{SDW}}$  is very small, while the iron isotope exponent  $\alpha_C = -\text{dln}T_C/\text{dln}M$  is about 0.35 (0.5 corresponds to the full isotope effect). Surprisingly, the iron isotope exchange shows the same effect on  $T_{\text{SDW}}$  as  $T_C$ . This indicates that electron–phonon interaction plays some role in the superconducting mechanism, but a simple electron–phonon coupling mechanism seems unlikely because a strong magnon–phonon coupling is included.

Recent inelastic neutron scattering measurements on  $\text{Ba}_{1-x}\text{K}_x\text{Fe}_2\text{As}_2$  ( $x = 0$  and 0.4) provide evidence for the presence of magnetic excitations<sup>10,11</sup>, suggesting that spin fluctuation may be important in the mechanism of superconductivity. However, phonons couple selectively to the spin system<sup>12</sup>. The structural transition from tetragonal to orthorhombic is driven by the antiferromagnetic SDW order<sup>13</sup>, and the antiferromagnetic SDW exists only in the orthorhombic structure<sup>14,15</sup>. The pressure coefficient of  $T_C$ ,  $\text{d}T_C/\text{d}P$ , changes from positive to negative with a crossover from orthorhombic to tetragonal symmetry for the superconducting phase<sup>14</sup>. The superconductivity and spin-density wave coexist in the orthorhombic structure<sup>14,16,17</sup>. These results indicate remarkable sensitivity of superconductivity and magnetism to the lattice.

We synthesized isotopically substituted polycrystalline samples with nominal compositions  $\text{SmFeAsO}_{1-x}\text{F}_x$  ( $x = 0, 0.15$ ) and  $\text{Ba}_{1-x}\text{K}_x\text{Fe}_2\text{As}_2$  ( $x = 0, 0.4$ ) by conventional solid-state reaction as described in refs 5 and 17, respectively. Figure 1 shows the Raman spectra for the samples  $\text{SmFeAsO}_{1-x}\text{F}_x$  by replacing  $^{16}\text{O}$  with  $^{18}\text{O}$ , and for the samples  $\text{SmFeAsO}_{1-x}\text{F}_x$  and  $\text{Ba}_{1-x}\text{K}_x\text{Fe}_2\text{As}_2$  by replacing  $^{56}\text{Fe}$  with the isotope  $^{54}\text{Fe}$ . The frequency shift of 4.2% and 4.5% for the  $E_g$  mode of oxygen<sup>18</sup> around  $420\text{ cm}^{-1}$  suggests about 71% substitution of  $^{18}\text{O}$  for  $^{16}\text{O}$  in the  $x = 0$  sample, and 77% for the  $x = 0.15$  sample. The Raman shift of about 1.7% we observed in the four samples for the  $B_{1g}$  mode of iron<sup>18</sup> indicates almost 100%  $^{54}\text{Fe}$  substitution for  $^{56}\text{Fe}$  for the two systems. These data are listed in Table 1.

The temperature dependence of resistivity  $\rho$  and its derivative  $\text{d}\rho/\text{d}T$  for typical samples  $\text{SmFeAsO}_{1-x}\text{F}_x$  on replacing  $^{16}\text{O}$  with the isotope  $^{18}\text{O}$  are shown in Fig. 2.  $T_C$  and  $T_{\text{SDW}}$  are listed in Table 2 for all samples from different batches. We calculated the isotope exponent  $\alpha_C$  for the superconducting transition using  $\alpha_C = -\text{dln}T_C/\text{dln}M$  to be  $-0.06(1)$ . To compare quantitatively the isotope effect on  $T_{\text{SDW}}$  with that on  $T_C$ , we also define an isotope



**Figure 1 | Raman spectra at room temperature for the  $\text{SmFeAsO}_{1-x}\text{F}_x$  ( $x = 0$  and 0.15) and  $\text{Ba}_{1-x}\text{K}_x\text{Fe}_2\text{As}_2$  ( $x = 0$  and 0.4) samples. **a**,  $\text{SmFeAsO}_{1-x}\text{F}_x$  ( $x = 0$  and 0.15) with  $^{18}\text{O}$  and  $^{16}\text{O}$ . **b**, As **a** with  $^{56}\text{Fe}$  and  $^{54}\text{Fe}$ . **c**,  $\text{Ba}_{1-x}\text{K}_x\text{Fe}_2\text{As}_2$  ( $x = 0$  and 0.4) with  $^{56}\text{Fe}$  and  $^{54}\text{Fe}$ . The mode around  $420\text{ cm}^{-1}$  is ascribed to the  $E_g$  mode of oxygen and the mode around  $210\text{ cm}^{-1}$  is assigned to the  $B_{1g}$  mode of Fe (ref. 18).**

<sup>1</sup>Hefei National Laboratory for Physical Sciences at Microscale and Department of Physics, University of Science and Technology of China, Hefei, Anhui 230026, China. <sup>2</sup>Beijing Synchrotron Radiation Facility, Institute of High Energy Physics, Chinese Academy of Sciences, Beijing 100049, China. <sup>3</sup>National Synchrotron Radiation Laboratory, University of Science and Technology of China, Hefei 230026, China.



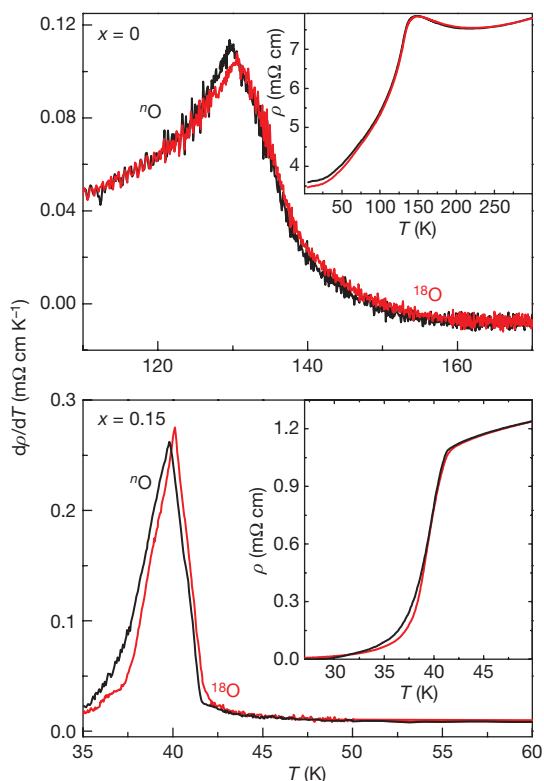
**Table 1 | Raman shifts indicate good extent of isotopic substitution**

| $\omega_{\text{O}} \text{ (cm}^{-1}\text{)}$                       | $^{16}\text{O}$  | $^{18}\text{O}$  | $\Delta\omega_{\text{O}}$  | $^{18}\text{O}$ (%)  |
|--|------------------|------------------|----------------------------|----------------------|
| SmFeAsO  | 418.6            | 401.2            | −17.4                      | 71                   |
| SmFeAsO <sub>0.85</sub> F <sub>0.15</sub>                          | 425.5            | 406.5            | −19                        | 77                   |
| $\omega_{\text{Fe}} \text{ (cm}^{-1}\text{)}$                      | $^{56}\text{Fe}$ | $^{54}\text{Fe}$ | $\Delta\omega_{\text{Fe}}$ | $^{54}\text{Fe}$ (%) |
| SmFeAsO  | 208.0            | 211.5            | 3.5                        | 92                   |
| SmFeAsO <sub>0.85</sub> F <sub>0.15</sub>                          | 210.6            | 214.4            | 3.8                        | 98                   |
| BaFe <sub>2</sub> As <sub>2</sub>                                  | 207.8            | 211.3            | 3.5                        | 92                   |
| Ba <sub>0.6</sub> K <sub>0.4</sub> Fe <sub>2</sub> As <sub>2</sub> | 204.8            | 208.2            | 3.4                        | 91                   |

Shown are the Raman shifts of the  $E_g$  mode of oxygen ( $\omega_{\text{O}}$ ) and the  $B_{1g}$  mode of iron ( $\omega_{\text{Fe}}$ ) in the SmFeAsO<sub>1-x</sub>F<sub>x</sub> ( $x = 0, 0.15$ ) samples with natural abundance oxygen  $^{16}\text{O}$  and  $^{18}\text{O}$ , and in the SmFeAsO<sub>1-x</sub>F<sub>x</sub> ( $x = 0, 0.15$ ) samples and Ba<sub>1-x</sub>K<sub>x</sub>Fe<sub>2</sub>As<sub>2</sub> ( $x = 0, 0.4$ ) samples with  $^{54}\text{Fe}$  and  $^{56}\text{Fe}$ , respectively. To calculate the  $^{18}\text{O}$  and  $^{54}\text{Fe}$  isotopic contents we assumed the Raman shift to be proportional to  $1 - (16/M)^{1/2}$  for the substitution of  $^{18}\text{O}$  for  $^{16}\text{O}$  and proportional to  $1 - (55.93/M)^{1/2}$  for the substitution of  $^{54}\text{Fe}$  for  $^{56}\text{Fe}$ , respectively. For the calculation, we used the following molar masses  $M$  for isotope exchange: 53.93 g for  $^{54}\text{Fe}$  and 55.93 g for  $^{56}\text{Fe}$ , and 16 g for  $^{16}\text{O}$  and 18 g for  $^{18}\text{O}$ .

exponent  $\alpha_{\text{SDW}} = -\text{dln}T_{\text{SDW}}/\text{dln}M$  for the SDW transition although no theory is yet established for the isotope effect on the magnetic phase transition. We obtained  $\alpha_{\text{SDW}} = -0.05(1)$ . These results indicate that the oxygen isotope effect on  $T_{\text{C}}$  and  $T_{\text{SDW}}$  is very small.

The temperature dependence of resistivity and its derivative for typical samples SmFeAsO<sub>1-x</sub>F<sub>x</sub> and Ba<sub>1-x</sub>K<sub>x</sub>Fe<sub>2</sub>As<sub>2</sub> when  $^{54}\text{Fe}$  is substituted for  $^{56}\text{Fe}$  is shown in Fig. 3. An increase in  $T_{\text{C}}$  is observed in the resistivity measurements, and  $\text{d}\rho/\text{d}T$  shows an increase of the SDW transition when  $^{54}\text{Fe}$  is substituted for  $^{56}\text{Fe}$ . The average results for several different samples are listed in Table 2. The average  $\alpha_{\text{SDW}}$  values for several samples of SmFeAsO and BaFe<sub>2</sub>As<sub>2</sub> are 0.39(2) and 0.36(2), and the average  $\alpha_{\text{C}}$  values for several samples of SmFeAsO<sub>0.85</sub>F<sub>0.15</sub> and Ba<sub>0.6</sub>K<sub>0.4</sub>Fe<sub>2</sub>As<sub>2</sub> are 0.34(3) and 0.37(3).



**Figure 2 | Temperature dependence of resistivity  $\rho$  and its derivative  $\text{d}\rho/\text{d}T$  for the SmFeAsO<sub>1-x</sub>F<sub>x</sub> samples with  $^{16}\text{O}$  and  $^{18}\text{O}$ .** The peak in the derivative is considered to be the transition temperature. The peak temperature in  $\text{d}\rho/\text{d}T$  corresponds to mid-transition temperature for the superconducting transition. An anomalous peak at about 147 K in resistivity is associated with the structural transition for the  $x = 0$  sample, and the peak temperature in  $\text{d}\rho/\text{d}T$  is very close to the SDW ordering temperature observed by neutron scattering<sup>23</sup>.

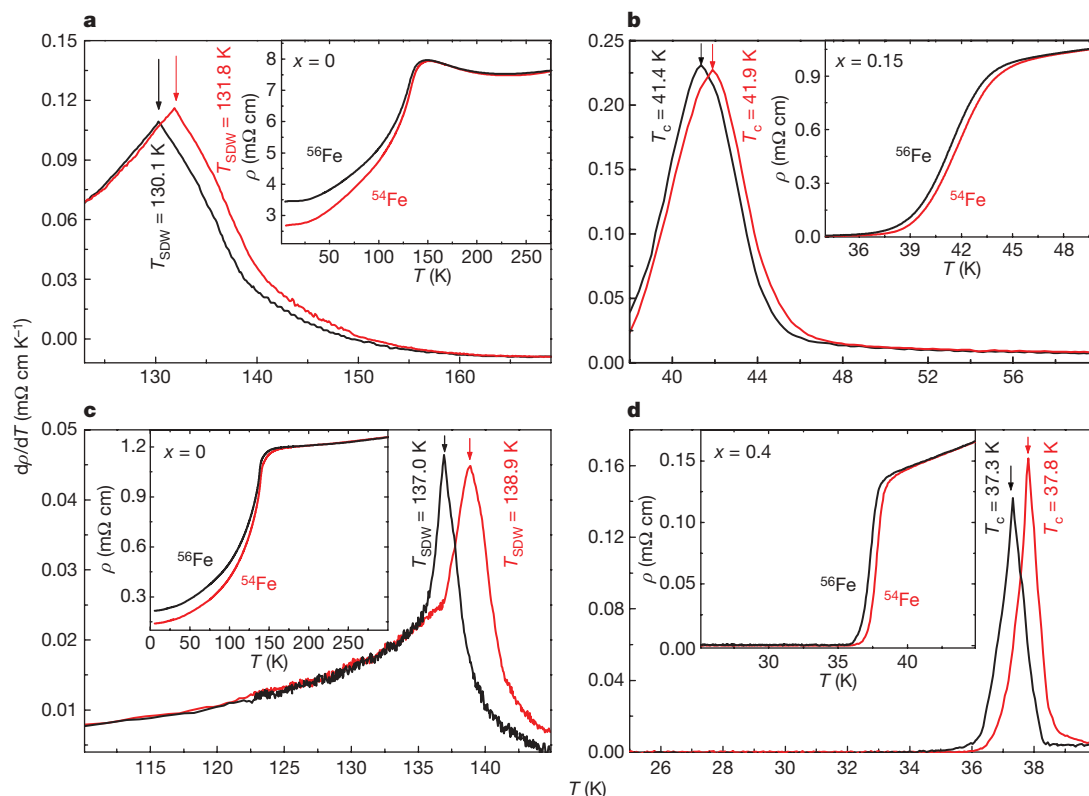
**Table 2 | Iron isotope effect on  $T_{\text{SDW}}$  and  $T_{\text{C}}$** 

| $T_{\text{SDW}}$   | $^{16}\text{O}$  | $^{18}\text{O}$  | $\Delta T_{\text{SDW}}$ | $\alpha_{\text{SDW}}$ |
|--|------------------|------------------|-------------------------|-----------------------|
| SmFeAsO  | 130.00(5)        | 130.80(5)        | 0.80(10)                | −0.07(1)              |
|  | 130.10(5)        | 130.60(5)        | 0.50(10)                | −0.04(1)              |
|  | 130.10(5)        | 130.60(5)        | 0.50(10)                | −0.04(1)              |
|  |                  |                  |                         | −0.05(1)<br>(average) |
| $T_{\text{SDW}}$   | $^{56}\text{Fe}$ | $^{54}\text{Fe}$ | $\Delta T_{\text{SDW}}$ | $\alpha_{\text{SDW}}$ |
| SmFeAsO  | 130.00(5)        | 131.90(5)        | 1.90(10)                | 0.41(2)               |
|  | 130.10(5)        | 131.80(5)        | 1.70(10)                | 0.37(2)               |
|  |                  |                  |                         | 0.39(2)<br>(average)  |
| BaFe <sub>2</sub> As <sub>2</sub>                                  | 137.00(5)        | 138.90(5)        | 1.90(10)                | 0.39(2)               |
|  | 137.10(5)        | 138.95(5)        | 1.85(10)                | 0.38(2)               |
|  | 137.10(5)        | 138.70(5)        | 1.60(10)                | 0.33(2)               |
|  |                  |                  |                         | 0.36(2)<br>(average)  |
| $T_{\text{C}}$   | $^{16}\text{O}$  | $^{18}\text{O}$  | $\Delta T_{\text{C}}$   | $\alpha_{\text{C}}$   |
| SmFeAsO <sub>0.85</sub> F <sub>0.15</sub>                          | 39.80(2)         | 40.10(2)         | 0.30(4)                 | −0.08(1)              |
|  | 41.70(2)         | 41.85(2)         | 0.15(4)                 | −0.04(1)              |
|  | 40.50(2)         | 40.85(2)         | 0.35(4)                 | −0.09(1)              |
|  | 41.50(2)         | 41.70(2)         | 0.20(4)                 | −0.05(1)              |
|  |                  |                  |                         | −0.06(1)<br>(average) |
| $T_{\text{C}}$   | $^{56}\text{Fe}$ | $^{54}\text{Fe}$ | $\Delta T_{\text{C}}$   | $\alpha_{\text{C}}$   |
| SmFeAsO <sub>0.85</sub> F <sub>0.15</sub>                          | 41.40(2)         | 41.90(2)         | 0.50(4)                 | 0.34(3)               |
|  | 40.90(2)         | 41.40(2)         | 0.50(4)                 | 0.34(3)               |
|  |                  |                  |                         | 0.34(3)<br>(average)  |
| Ba <sub>0.6</sub> K <sub>0.4</sub> Fe <sub>2</sub> As <sub>2</sub> | 37.30(2)         | 37.80(2)         | 0.50(4)                 | 0.38(3)               |
|  | 37.30(2)         | 37.82(2)         | 0.52(4)                 | 0.39(3)               |
|  | 37.40(2)         | 37.86(2)         | 0.46(4)                 | 0.35(3)               |
|  |                  |                  |                         | 0.37(3)<br>(average)  |

Shown are  $T_{\text{SDW}}$  and  $T_{\text{C}}$ , and their shifts ( $\Delta T_{\text{SDW}}$  and  $\Delta T_{\text{C}}$ ) and  $\alpha_{\text{SDW}}$  and  $\alpha_{\text{C}}$  for the SmFeAsO<sub>1-x</sub>F<sub>x</sub> samples with natural abundance oxygen  $^{16}\text{O}$  and  $^{18}\text{O}$ , and  $^{56}\text{Fe}$  and  $^{54}\text{Fe}$ , and for the Ba<sub>1-x</sub>K<sub>x</sub>Fe<sub>2</sub>As<sub>2</sub> samples with  $^{54}\text{Fe}$  and  $^{56}\text{Fe}$ , respectively. We calculated isotope exponents using  $\alpha = -\text{dln}T/\text{dln}M$ , where  $M$  is the atomic mass. To guarantee that the intrinsic isotope effect is what is being measured, isotope back-exchange is usually used: the transition should return to that before isotope exchange. In our case, there is no way of carrying out the isotope back-exchange experiment. To obtain the intrinsic isotope effect, we prepared several batches of samples to check that the results are reproducible from sample to sample. Although  $T_{\text{C}}$  varies slightly from batch to batch for SmFeAsO<sub>0.85</sub>F<sub>0.15</sub>, the  $T_{\text{C}}$  shift caused by isotope exchange is nearly the same for all batches.  $T_{\text{C}}$  for the sample Ba<sub>0.6</sub>K<sub>0.4</sub>Fe<sub>2</sub>As<sub>2</sub> is very stable and nearly independent of batches. In the calculation the molar masses used are the same as used in Table 1 except for  $^{18}\text{O}$ , reduced to 17.42 g and 17.54 g because the isotope replacement was only 71% and 77%. The error bar for the temperature determination is equal to the step temperature in the temperature sweep: 0.02 K for  $T_{\text{C}}$  and 0.05 K for  $T_{\text{SDW}}$  in the resistivity measurements, respectively.

These values are comparable to 0.5 for the full isotope effect in the framework of BCS theory and indicate a strong iron isotope effect on  $T_{\text{C}}$  and  $T_{\text{SDW}}$ . This implies that the electron–phonon interaction should play an important role for the superconducting mechanism. We note that  $\alpha_{\text{C}}$  and  $\alpha_{\text{SDW}}$  for  $T_{\text{C}}$  and  $T_{\text{SDW}}$  are almost the same for the two systems, and much larger than the oxygen isotope exponents.

Isotope effect studies require well-characterized samples with reproducible crystal chemistry properties. We found that  $T_{\text{SDW}}$  is insensitive to the sample processing for the parent compounds. However, because the F content is not easy to control,  $T_{\text{C}}$  for the SmFeAsO<sub>0.85</sub>F<sub>0.15</sub> sample is sensitive to the sample processing. A detailed description of the synthesis procedure we used to ensure the same F content is given in the Supplementary Information. No difference in the lattice constants (see Supplementary Fig. S1) provides strong evidence for the same F content for isotope exchange. To confirm that the observed results are intrinsic instead of an impurity effect, we checked whether  $T_{\text{C}}$  and  $T_{\text{SDW}}$  differ for the Ba<sub>1-x</sub>K<sub>x</sub>Fe<sub>2</sub>As<sub>2</sub> samples using natural abundance iron ( $^{56}\text{Fe}$ ) with 99.9% purity and  $^{54}\text{Fe}$  with 99.78% purity.  $T_{\text{SDW}}$  and  $T_{\text{C}}$  are nearly the same for the samples with  $^{56}\text{Fe}$  and  $^{54}\text{Fe}$  (see Supplementary Fig. S3). We synthesized the Ba<sub>0.6</sub>K<sub>0.4</sub>Fe<sub>2</sub>As<sub>2</sub> samples



**Figure 3 | Temperature dependence of  $\rho$  and  $d\rho/dT$  for the  $\text{SmFeAsO}_{1-x}\text{F}_x$  and  $\text{Ba}_{1-x}\text{K}_x\text{Fe}_2\text{As}_2$  samples isotopically substituted with  $^{56}\text{Fe}$  and  $^{54}\text{Fe}$ .** **a**,  $\text{SmFeAsO}_{1-x}\text{F}_x$  with  $x = 0$ . **b**,  $\text{SmFeAsO}_{1-x}\text{F}_x$  with  $x = 0.15$ . **c**,  $\text{Ba}_{1-x}\text{K}_x\text{Fe}_2\text{As}_2$  with  $x = 0$ . **d**,  $\text{Ba}_{1-x}\text{K}_x\text{Fe}_2\text{As}_2$  with  $x = 0.4$ . To determine the transition temperature accurately, we took the derivative of the resistivity, and the peak in the derivative is considered the transition temperature. The peak temperature in  $d\rho/dT$  for the SDW is very close to the ordering temperature observed by neutron scattering<sup>23</sup>. We deduced  $\alpha_C$  and  $\alpha_{\text{SDW}}$  to be 0.34(3) and 0.37(2) for the  $\text{SmFeAsO}_{1-x}\text{F}_x$  samples with  $x = 0.15$  and 0, respectively. For a typical sample of  $\text{BaFe}_2\text{As}_2$ , the resistivity anomaly around 145 K arises from the structural and SDW transition<sup>24</sup>. For a typical superconducting sample of  $\text{Ba}_{1-x}\text{K}_x\text{Fe}_2\text{As}_2$  with  $x = 0.4$ , the onset transition occurs around 39 K. We found  $\alpha_C$  and  $\alpha_{\text{SDW}}$  to be 0.38(3) and

0.39(2) for the  $\text{Ba}_{1-x}\text{K}_x\text{Fe}_2\text{As}_2$  samples with  $x = 0.4$  and 0, respectively. To confirm the results from resistivity measurements, we measured the susceptibility for the same samples (see Supplementary Fig. S2). The shift of  $T_C$  and  $T_{\text{SDW}}$  determined from susceptibility measurements due to iron isotope exchange is consistent with that obtained from the resistivity measurements. A remarkable difference in resistivity is observed below the SDW transition between the  $x = 0$  samples with  $^{56}\text{Fe}$  and  $^{54}\text{Fe}$  for the  $\text{SmFeAsO}_{1-x}\text{F}_x$  and  $\text{Ba}_{1-x}\text{K}_x\text{Fe}_2\text{As}_2$  systems. As shown in the insets of Fig. 3a and c, the resistivity decreases by about 28% for the  $\text{SmFeAsO}$  sample and by about 35% for the  $\text{BaFe}_2\text{As}_2$  sample at 10 K owing to iron isotope exchange, suggesting that iron isotope exchange has a strong effect on the SDW state.

using  $^{56}\text{Fe}$  with 98% and 99.9% purity to determine the effect of impurity on  $T_C$ . The difference of  $T_C$  for the two samples is 0.07 K (see Supplementary Fig. S4), indicating that the effect of impurity on  $T_C$  is very small, and does not affect the intrinsic isotope effect observed in Table 2.

We emphasize that iron isotope exchange has a strong effect on the SDW state. Substitution of  $^{54}\text{Fe}$  for  $^{56}\text{Fe}$  leads to a remarkable decrease in resistivity below the SDW ordering temperatures with a large  $\alpha_{\text{SDW}}$  for the two systems, suggesting a strong magnon–phonon coupling. A giant oxygen isotope effect has been observed in magnetoresistive  $\text{La}_{1-x}\text{Ca}_x\text{MnO}_{3+y}$  and the isotope exponent  $\alpha_{\text{FM}}$  for ferromagnetic transition is as high as 0.85 (ref. 19). This large isotope shift is believed to arise from coupling of the charge carriers to Jahn–Teller lattice distortions<sup>19</sup>. In pnictide superconductors, the strong sensitivity of superconductivity and magnetism to the lattice may be responsible for the large isotope effect. These results indicate that the electron–phonon interaction is important in the superconducting mechanism, but the strong magnon–phonon coupling must also be considered.

The iron isotope effect on  $T_{\text{SDW}}$  and  $T_C$  is much larger than the oxygen isotope effect in pnictide superconductors. The reason could be that the iron–arsenide plane is the conducting layer and thus responsible for the superconductivity, and the SDW ordering originates from the Fe moment. For the  $\text{MgB}_2$  superconductor, no magnetic correlation is included and the superconductivity can be

understood within BCS theory with  $\alpha_C = 0.32$  (ref. 20). In the copper oxides, the isotope effect on  $T_C$  is sensitive to doping level. The effect is vanishing at optimum doping, but increases systematically with decreasing doping level to a maximum at the border to the antiferromagnetic state<sup>21,22</sup>. It thus seems that the isotope effect is related to magnetic fluctuation. Such unconventional isotope effects demonstrate that the electron–phonon interaction is also important in the physics of copper oxides. Sorting out the interplay between the lattice and magnetic degrees of freedom is a key challenge for our understanding the mechanism of high- $T_C$  superconductivity.

Received 4 November 2008; accepted 13 March 2009.

- Kamihara, Y. *et al.* Iron-based layered superconductor  $\text{LaO}_{1-x}\text{F}_x\text{FeAs}$  ( $x = 0.05\text{--}0.12$ ) with  $T_C = 26$  K. *J. Am. Chem. Soc.* **130**, 3296–3297 (2008).
- Chen, X. H. *et al.* Superconductivity at 43 K in  $\text{SmFeAsO}_{1-x}\text{F}_x$ . *Nature* **453**, 761–762 (2008).
- Chen, G. F. *et al.* Superconductivity at 41 K and its competition with spin-density-wave instability in layered  $\text{CeO}_{1-x}\text{F}_x\text{FeAs}$ . *Phys. Rev. Lett.* **100**, 247002 (2008).
- Ren, Z. A. *et al.* Superconductivity in iron-based F-doped layered quaternary compound  $\text{NdO}_{1-x}\text{F}_x\text{FeAs}$ . *Europhys. Lett.* **82**, 57002 (2008).
- Liu, R. H. *et al.* Anomalous transport properties and phase diagram of the FeAs-based  $\text{SmFeAsO}_{1-x}\text{F}_x$  superconductors. *Phys. Rev. Lett.* **101**, 087001 (2008).
- Boeri, L. *et al.* Is  $\text{LaFeAsO}_{1-x}\text{F}_x$  an electron–phonon superconductor? *Phys. Rev. Lett.* **101**, 026403 (2008).
- Cao, C. *et al.* Proximity of antiferromagnetism and superconductivity in  $\text{LaFeAsO}_{1-x}\text{F}_x$ : effective Hamiltonian from ab initio studies. *Phys. Rev. B* **77**, 220506(R) (2008).



8. Dai, X. *et al.* Even parity, orbital singlet, and spin triplet pairing for superconducting  $\text{LaFeAsO}_{1-x}\text{F}_x$ . *Phys. Rev. Lett.* **101**, 057008 (2008).
9. Ma, F. *et al.* Iron-based layered compound  $\text{LaFeAsO}$  is an antiferromagnetic semimetal. *Phys. Rev. B* **78**, 033111 (2008).
10. Ewings, R. A. *et al.* High energy spin excitations in  $\text{BaFe}_2\text{As}_2$ . *Phys. Rev. B* **78**, 220501(R) (2008).
11. Christianson, A. D. *et al.* Resonant spin excitation in the high temperature superconductor  $\text{Ba}_{0.6}\text{K}_{0.4}\text{Fe}_2\text{As}_2$ . *Nature* **456**, 930–932 (2008).
12. Zbiri, M. *et al.* Combined ab initio lattice dynamics simulations and improved inelastic neutron scattering spectra for studying phonons in  $\text{BaFe}_2\text{As}_2$ : effect of structural phase transition, structural relaxation and magnetic ordering. Preprint at <<http://arxiv.org/abs/0807.4429>> (2008).
13. Fang, C. *et al.* Theory of electron nematic order in  $\text{LaOFeAs}$ . *Phys. Rev. B* **77**, 224509 (2008).
14. Margadonna, S. *et al.* Crystal structure and phase transitions across the metal-superconductor boundary in the  $\text{SmFeAsO}_{1-x}\text{F}_x$  ( $0 \leq x \leq 0.20$ ) family. *Phys. Rev. B* **79**, 014503 (2009).
15. Zhao, J. *et al.* Structural and magnetic phase diagram of  $\text{CeFeAsO}_{1-x}\text{F}_x$  and its relation to high-temperature superconductivity. *Nature Mater.* **7**, 953–959 (2008).
16. Drew, A. J. *et al.* Coexistence of static magnetism and superconductivity in  $\text{SmFeAsO}_{1-x}\text{F}_x$  as revealed by muon spin rotation. *Nature Mater.* **8**, 310–314 (2009).
17. Chen, H. *et al.* Coexistence of the spin-density-wave and superconductivity in the  $(\text{Ba,K})\text{Fe}_2\text{As}_2$ . *Europhys. Lett.* **85**, 17006 (2009).
18. Zhao, S. C. *et al.* Raman spectra in iron-based quaternary  $\text{CeO}_{1-x}\text{F}_x\text{FeAs}$  and  $\text{LaO}_{1-x}\text{F}_x\text{FeAs}$ . *Supercond. Sci. Technol.* **22**, 015017 (2009).
19. Zhao, G. M. *et al.* A giant oxygen isotope shift in magnetoresistive perovskite  $\text{La}_{1-x}\text{Ca}_x\text{MnO}_{3+y}$ . *Nature* **381**, 676–678 (1996).
20. Hinks, D. G., Claus, H. & Jorgensen, J. D. The complex nature of superconductivity in  $\text{MgB}_2$  as revealed by the reduced total isotope effect. *Nature* **441**, 457–460 (2001).
21. Franck, J. P. in *Physical Properties of High Temperature Superconductors* (ed. Ginsberg, D. M.) 189 (World Scientific, 1994).
22. Keller, H. in *Structure and Bonding* (eds Müller, K. A. & Bussmann-Holder, A.) Vol. 114, 143 (Springer, 2005).
23. Cruz, C. *et al.* Magnetic order close to superconductivity in the iron-based layered  $\text{LaO}_{1-x}\text{F}_x\text{FeAs}$  systems. *Nature* **453**, 899–902 (2008).
24. Huang, Q. *et al.* Magnetic order in  $\text{BaFe}_2\text{As}_2$ , the parent compound of the FeAs based superconductors in a new structural family. *Phys. Rev. Lett.* **101**, 257003 (2008).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This work is supported by the Nature Science Foundation of China, and by the Ministry of Science and Technology of China and Chinese Academy of Sciences. We acknowledge Z. X. Shen for discussion and encouragement, and D. L. Feng and S. Y. Li for discussions.

**Author Contributions** X.H.C. designed and coordinated the whole experiment, and analysed the data and wrote the paper. R.H.L. and T.W. performed the main experiments, including sample preparation and analysed the data. G.W., X.F.W. and B.C.S. synthesized the samples. H.C. and Y.L.X. partially measured the resistivity. J.J.Y. measured the susceptibility. Y.J.Y. and Q.J.L. did X-ray powder diffraction measurements. W.S.C. and Z.Y.W. provided the iron isotope  $^{54}\text{Fe}$ .

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to X.H.C. ([chenhx@ustc.edu.cn](mailto:chenhx@ustc.edu.cn)).

## LETTERS

# Force-induced activation of covalent bonds in mechanoresponsive polymeric materials

Douglas A. Davis<sup>1</sup>, Andrew Hamilton<sup>2</sup>, Jinglei Yang<sup>3†</sup>, Lee D. Cremer<sup>1</sup>, Dara Van Gough<sup>4</sup>, Stephanie L. Potisek<sup>1</sup>, Mitchell T. Ong<sup>1</sup>, Paul V. Braun<sup>1,3,4</sup>, Todd J. Martínez<sup>1,3†</sup>, Scott R. White<sup>3,5</sup>, Jeffrey S. Moore<sup>1,3,4</sup> & Nancy R. Sottos<sup>3,4</sup>

Mechanochemical transduction enables an extraordinary range of physiological processes such as the sense of touch, hearing, balance, muscle contraction, and the growth and remodelling of tissue and bone<sup>1–6</sup>. Although biology is replete with materials systems that actively and functionally respond to mechanical stimuli, the default mechanochemical reaction of bulk polymers to large external stress is the unselective scission of covalent bonds, resulting in damage or failure<sup>7</sup>. An alternative to this degradation process is the rational molecular design of synthetic materials such that mechanical stress favourably alters material properties. A few mechanosensitive polymers with this property have been developed<sup>8–14</sup>; but their active response is mediated through non-covalent processes, which may limit the extent to which properties can be modified and the long-term stability in structural materials. Previously, we have shown with dissolved polymer strands incorporating mechanically sensitive chemical groups—so-called mechanophores—that the directional nature of mechanical forces can selectively break and re-form covalent bonds<sup>15,16</sup>. We now demonstrate that such force-induced covalent-bond activation can also be realized with mechanophore-linked elastomeric and glassy polymers, by using a mechanophore that changes colour as it undergoes a reversible electrocyclic ring-opening reaction under tensile stress and thus allows us to directly and locally visualize the mechanochemical reaction. We find that pronounced changes in colour and fluorescence emerge with the accumulation of plastic deformation, indicating that in these polymeric materials the transduction of mechanical force into the ring-opening reaction is an activated process. We anticipate that force activation of covalent bonds can serve as a general strategy for the development of new mechanophore building blocks that impart polymeric materials with desirable functionalities ranging from damage sensing to fully regenerative self-healing.

We created mechanoresponsive synthetic polymeric materials by directly linking force-activated mechanophores into the polymer chains of bulk polymers or by using the mechanophores as cross-links (Fig. 1). The mechanophore motif must efficiently transfer external force to a relatively small number of specific bonds in the bulk polymer, with this mechanochemical activation depending critically on the molecular structure of the mechanophore and the attachment points to the polymer chains. To demonstrate directly that external forces can locally activate a mechanophore linked into elastomeric or glassy polymer solids, we first selected a mechanophore with the potential to undergo a force-induced, 6- $\pi$  electrocyclic ring-opening reaction<sup>17</sup> that is accompanied by a colour change (Fig. 1a). The selected molecule is well-characterized and exhibits both thermo- and photochromic responses<sup>18</sup> that reversibly transform a closed,

colourless spiropyran form to a highly coloured, planar merocyanine structure through rupture of the spiro carbon-oxygen (C–O) bond. The increased conjugation of the merocyanine form shifts the absorption to longer wavelengths, thus giving rise to visible colour and fluorescence. The spiro C–O bond is the critical point for mechanically induced activation, and there are many possible attachment points on the spiropyran that differ in their ability to transmit force efficiently across the junction (Fig. 1a). Simple mechanical models indicate that attachment on opposing sides of the spiro-junction, specifically positions 5' or 6' of the indole side and positions 7 or 8 of the benzopyran side, preferentially stresses the spiro C–O bond over the spiro C–C bond.

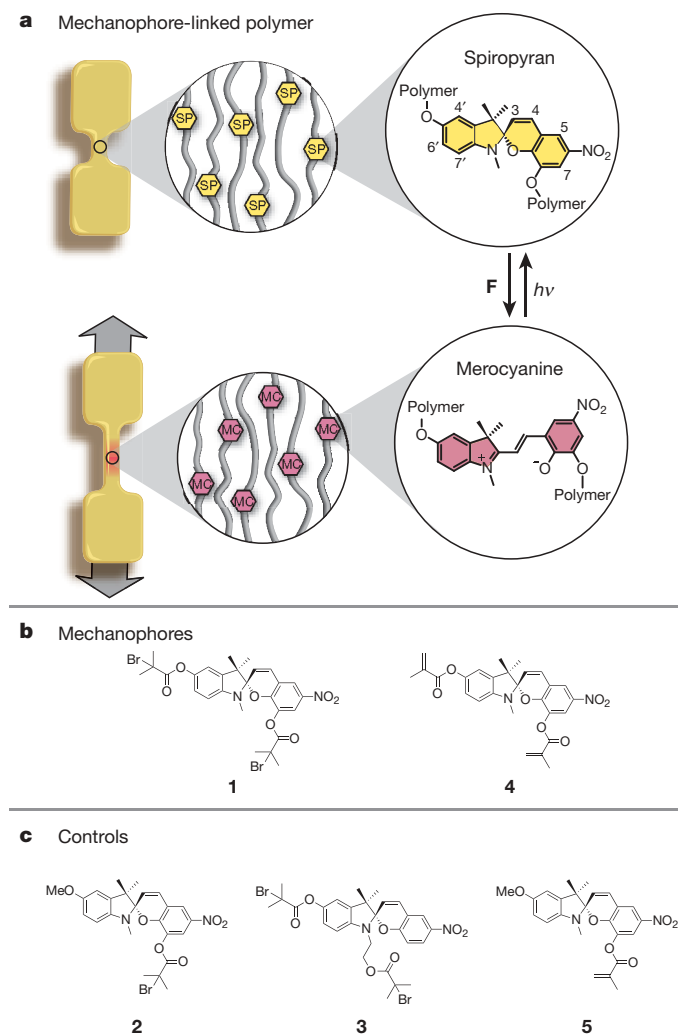
We selected positions 5' and 8 as the attachment points and functionalized them with  $\alpha$ -bromo or methacryloyl esters for subsequent polymerization by single-electron-transfer living radical polymerization<sup>19</sup> or suspension radical polymerization, respectively (Fig. 1b, structures 1 and 4). We prepared control structures to determine whether the observed colour change was dependent on linking through the spiro-junction or was the result of other factors, such as localized heating or radical generation from polymer chain cleavage during mechanical deformation<sup>20</sup>. The monofunctional control lacked the polymer chain attached to the indole portion (Fig. 1c, structures 2 and 5), precluding the transfer of any substantial force into the spiropyran. A difunctional control (Fig. 1c, structure 3) with a structure closer to the purported mechanophore was prepared by placing the spiropyran in the centre of the polymer chain, but with polymer chains linked only to one side of the spiro-junction. This placement allowed the transfer of force to the spiropyran, but not to the spiro-fused C–O bond specifically.

We modelled the effect of external force on the mechanophore (Fig. 2a) with both first-principles steered molecular dynamics<sup>21–23</sup> and constrained optimization (COGEF)<sup>24</sup> simulations, and using truncated (**t**) as well as extended (**e**) models of the mechanophore unit 1 (insets to Fig. 2c and d, see also Supplementary Figs S6 and S7). The results of dynamics simulations using density functional theory (DFT) are shown in Fig. 2b for mechanophore model **1t** and difunctional control model **3t**. The C–O distance at the spiro bond is plotted as a function of time for a variety of applied forces, ranging from 2 to 3 nN. For **1t**, scission occurred exclusively at the C–O bond for all applied forces investigated within this range (Supplementary Video S2). For **3t**, no C–O bond rupture was observed on the indicated timescale (Supplementary Video S3).

The origin of this behaviour was further explored using the COGEF procedure at the DFT level (Fig. 2c). For **1t**, the potential energy of the molecule rose as the distance between the attachment points increased, leading to a reaction activation barrier of 2.3 eV at an

<sup>1</sup>Department of Chemistry, <sup>2</sup>Department of Mechanical Science and Engineering, <sup>3</sup>The Beckman Institute, <sup>4</sup>Department of Materials Science and Engineering, <sup>5</sup>Department of Aerospace Engineering, University of Illinois at Urbana-Champaign, Illinois 61801, USA. †Present addresses: School of Mechanical and Aerospace Engineering, Nanyang Technological University, Singapore (J.Y.); Department of Chemistry, Stanford University, Stanford, California, USA (T.J.M.).





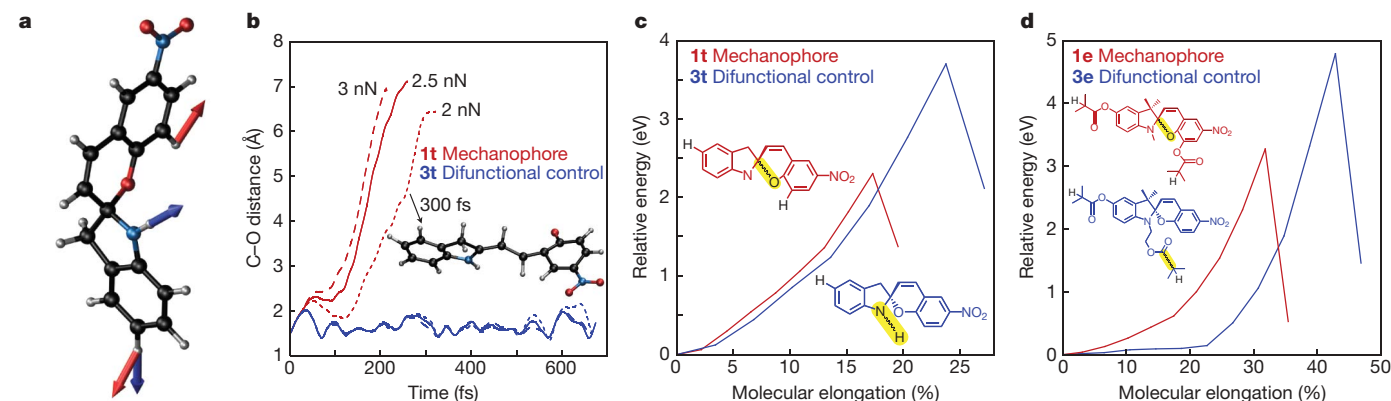
elongation of 17%; further elongation of the molecule to 20% ruptured the spiro C–O bond. In contrast, **3t** showed no C–O bond rupture at 20% elongation and the activation barrier increased to 3.7 eV at an elongation of 24%; further elongation finally caused rupture of the N–H bond. In the case of the extended models, the DFT COGEF

**Figure 1 | Chemical structures and bulk polymeric samples.** **a**, Schematic diagram of ‘dog bone’ specimens prepared from linear 80 kDa PMA. Upon application of tensile force, a hypothesized conversion between the colourless spiropyran and coloured merocyanine forms of the mechanophore occurs. Exposure to visible light reverses the conversion back to the original spiropyran form. **b**, The polymer connectivity of spiropyran **1** and **4** is across the spiro-junction, and therefore these materials are expected to be mechanically active. **c**, Materials derived from spiropyran **2**, **3** and **5** serve as controls, since the polymer connectivity does not include the spiro-junction. Spiropyran functionalized with  $\alpha$ -bromo esters were used to initiate single-electron-transfer living radical polymerization and incorporate a single spiropyran in the middle (**1**, **3**) or at the end (**2**) of the polymer chain. Spiropyran functionalized with methacryloyl esters (**4**, **5**) were copolymerized with methyl methacrylate in aqueous suspension to form PMMA beads incorporating spiropyran at cross-linked junctions (**4**) or as pendant groups (**5**).

energy profiles for **1e** (Fig. 2d) again showed spiro C–O bond rupture, whereas the difunctional control molecule **3e** underwent a bond rupture at one of the side chains that does not lead to a colour change. Rupture of this side chain required more energy than that needed to activate **1e**. Thus, the COGEF calculations predicted selective spiro C–O bond rupture for the **1t** as well as the **1e** models. Semi-empirical calculations were in good agreement with DFT for both the truncated and extended models (Supplementary Figs S9 and S12), justifying their use in longer-lasting dynamics simulations to explore the effects of lower applied forces.

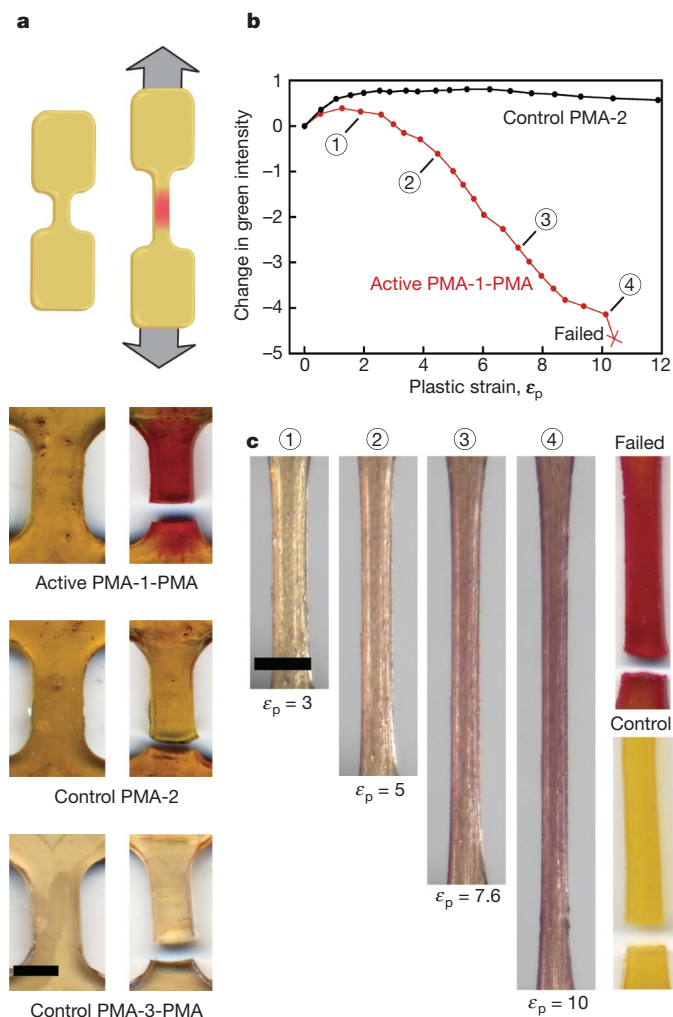
Experimentally, mechanochemical transduction was demonstrated for an elastomeric mechanophore-linked poly(methyl acrylate) (PMA) and a glassy mechanophore cross-linked poly(methyl methacrylate) (PMMA). Elastomeric materials consisted of **1**, **2** or **3** polymerized into PMA, resulting in a single mechanophore per chain. Polymer samples were heat moulded into a ‘dog bone’ shape, which was loaded either monotonically or cyclically under displacement control while monitoring the applied stress (Fig. 3a). Glassy materials consisted of solutions of **4** and **5** in a methyl methacrylate suspension copolymerized using benzoyl peroxide and *N,N*-dimethylaniline to give 100–500  $\mu\text{m}$  diameter beads (Fig. 4a). These PMMA beads were compressed with a stepper actuator coupled to a load cell, which allowed for strain rate control while monitoring the resulting stress.

The active mechanophore-linked PMA (PMA-1-PMA), the monofunctional control (PMA-2), and the difunctional control (PMA-3-PMA) specimens were tested to failure under monotonic tensile loading. An intense colour change occurred in the active specimen (the absorption band is in the 550–590 nm range; Supplementary



**Figure 2 | First-principles dynamics and constrained optimization models of mechanical activation.** **a**, Truncated model of spiropyran showing the pulling directions (mechanophore **1t**, red arrows; control **3t**, blue arrows). The extended model includes ester chains in place of the H atoms (see inset of **d**). **b**, Steered molecular dynamics for the truncated mechanophore and control. The C–O bond distance is plotted versus time at 2, 2.5 and 3 nN, as shown. **c**, Potential energy versus percentage elongation for the truncated models (**1t** and **3t**) calculated at the DFT level. Elongation, the distance

between the pulled H atoms, is increased systematically and all other geometric coordinates are optimized at each point. Bond rupture occurs at the highlighted spiro C–O and N–H bonds for **1t** and **3t**, respectively. **d**, Potential energy for the extended models (**1e** and **3e**) calculated with DFT. Elongation is the distance between the terminal H (instead of Br in **1** and **3**) atoms on the side chains. Rupture for **1e** and **3e** occurs at the highlighted C–O spiro and C–C ester bonds, respectively.



**Figure 3 | Mechanochromic response of mechanophore-linked PMA elastomer under tensile loading.** **a**, Schematic of tensile loading direction. Optical images of active PMA-1-PMA, monofunctional control PMA-2, and difunctional control PMA-3-PMA specimens before testing and after failure in monotonic tension. **b**, Accumulation of plastic (unrecovered) strain and relative change in green intensity for active PMA-1-PMA and monofunctional PMA-2 control samples after each loading cycle in a fatigue test. **c**, Optical images of an active PMA-1-PMA specimen acquired after the loading cycles indicated in **b** and just after failure. For comparison, a failed monofunctional PMA-2 control sample subjected to the same cyclic loading regimen is included (lower right). Scale bars in **a** and **c** are 2 mm.

Fig. S17), whereas no colour change was observed in the monofunctional or difunctional controls (Fig. 3a). The mechanically induced colour change could be reversed after approximately 6 h of exposure to fluorescent room light.

We further evaluated the evolution of the colour change in the active PMA-1-PMA by a cyclic loading regime, in which increasing levels of strain were applied with each cycle (Supplementary Figs S3 and S4) causing plastic (unrecovered) strain to accumulate during the testing time frame. Digital images of the gauge section were acquired after each cycle, and the change in the RGB (red, green, blue) colour ratios was analysed as described in Supplementary Fig. S14.

The representative images of the active PMA-1-PMA samples undergoing cyclic loading in Fig. 3c reveal the emergence of red colour with increasing plastic deformation. The corresponding relative change in the green channel intensity is plotted in Fig. 3b as a function of the accumulated strain at the end of each loading cycle. The active PMA-1-PMA shows a nearly monotonic reduction in the relative green intensity after reaching a plastic strain level of 200%. In contrast, the monofunctional control PMA-2 exhibited no change in the green

intensity level throughout the loading regime. After reaching a strain level of just over 1,000%, the active PMA-1-PMA failed, leaving an intense red colour throughout the entire gauge section. An image of the failed monofunctional control PMA-2 sample is also included in Fig. 3c for comparison. The lack of activation in the control samples provides strong evidence that the colour change is induced solely by the application of mechanical force, rather than by temperature change or photolytic activation (Supplementary Table S1).

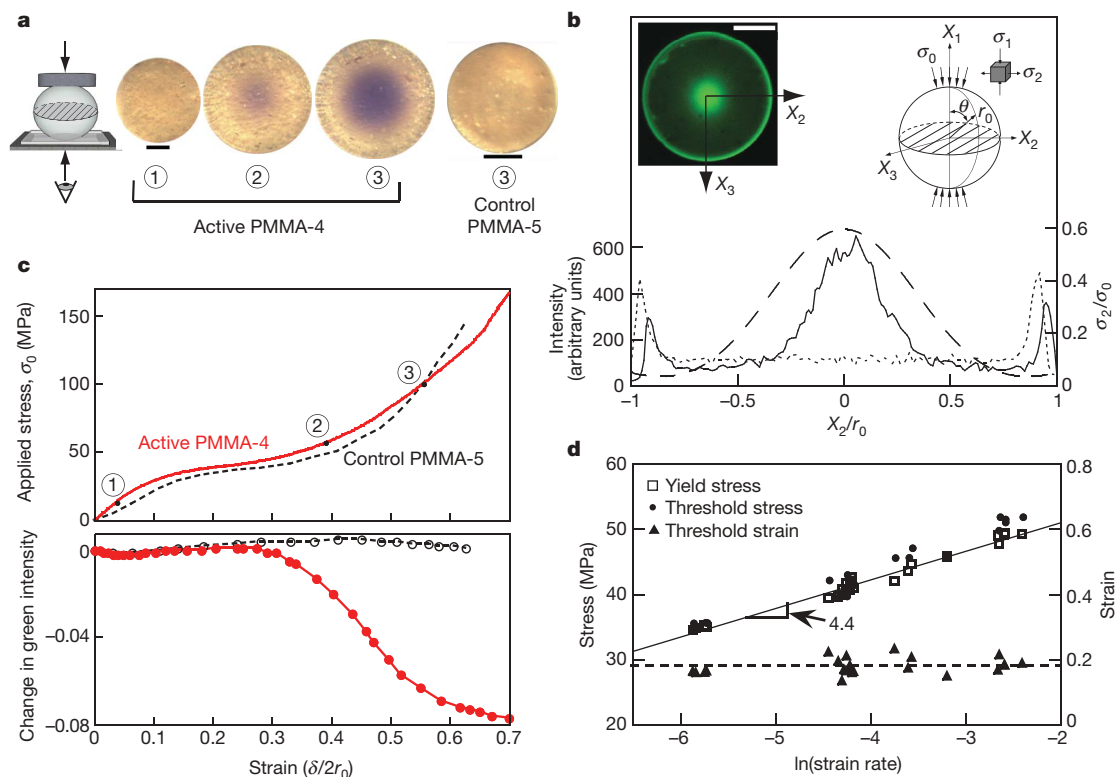
We adapted a compressive loading configuration (Fig. 4a) for the glassy, mechanophore cross-linked PMMA-4 and corresponding monofunctional PMMA-5 control beads. Although compressive stresses ( $\sigma_1$ ) develop in the bead parallel to the loading direction ( $X_1$ ), significant tensile stresses ( $\sigma_2$ ) are induced perpendicular to the loading direction ( $X_2$ ). Elastic predictions<sup>25,26</sup> of the  $\sigma_2$  stress field on the equatorial plane of a compressed bead are plotted along the  $X_2$  axis in Fig. 4b, and a more detailed stress analysis is included in the Supplementary Information. The maximum tensile stress develops in the centre of the bead ( $r=0$ ) and reaches a value of 0.6 of the magnitude of the applied compressive stress ( $\sigma_o$ ). Hence, compressive loading generates tensile stresses that are expected to activate mechanophore 4 in this configuration. For this specimen geometry, we characterized mechanochemical activation by analysis of the relative change in RGB intensity during loading (Supplementary Video S7) and the relative change in fluorescence after testing.

The evolution of a mechanochemically induced colour change in the active cross-linked PMMA-4 bead is shown in Fig. 4a and c for a representative compression test. The colour change emerges in the centre of the bead at a strain level just beyond the yield point. The corresponding fluorescence intensity, as measured by confocal laser scanning microscopy, is plotted across the equatorial plane in Fig. 4b. The maximum fluorescence intensity at the bead centre correlates with the predicted tensile stress distribution (just before yield) in the same plane. No detectable change in colour or significant fluorescence was observed for the monofunctional control PMMA-5 beads (Supplementary Fig. S18). As with the mechanophore-linked PMA, the evolution of colour in the active cross-linked bead corresponds to a reduction of the relative green channel intensity, and the colour (and fluorescence) intensifies with accumulation of plastic strain. The colour change in the glassy PMMA is more stable than in PMA and remains after several weeks of exposure to fluorescent room light.

To explore further the relationship of colour change with the onset of yielding, we characterized the threshold stress and strain levels required to produce colour change in the active cross-linked PMMA-4 beads over a range of strain rates. As shown in Fig. 4d, the yield stress and threshold stress both increased linearly with the natural log of the strain rate, while the threshold strain remained nearly constant. This relationship between threshold stress, yield stress and strain rate indicates that the mechanochemical reaction in the bulk polymer may be a strain-activated rate process<sup>27</sup>. With plastic flow above a strain threshold, activation of the mechanophore-linked polymer is achieved, and a colour signature is obtained.

We have shown that routinely encountered loads and strain rates are sufficient to trigger covalent bond changes with appropriately designed mechanophores. The spiropyran mechanophore can function as a molecular force sensor, providing visible detection and mapping of mechanical stresses within bulk polymeric materials. We expect that the spiropyran mechanophore will serve as a molecular probe to aid in understanding the effects of stress and damage on polymeric materials, providing an opportunity for assessment, modification and improvement before catastrophic failure. Overall, the mechanophore motif is general and capable of translating many other desirable responses including polymerization, cross-linking or even strain-activated mechanical fuses. With a deeper understanding of mechanophore design and efficient chemical response pathways, we envision new classes of dynamically responsive polymers that locally remodel, reorganize or even regenerate upon mechanical stimulation.





**Figure 4 | Mechanochromic response of glassy mechanophore cross-linked PMMA beads under diametral compression.** **a**, Schematic of compression loading; deformed beads are imaged from bottom. Optical images as indicated on the stress–strain curve in **c**. Scale bars 50  $\mu\text{m}$ . **b**, Measured fluorescence intensity in the equatorial plane for a bead deformed just past the yield point (left inset, scale bar 50  $\mu\text{m}$ ). Initial signature (dotted line) and fluorescence intensity after deformation (solid line) plotted along the  $X_2$  axis. Elastic prediction (dashed line) of the

normalized transverse stress field,  $\sigma_2/\sigma_0$  ( $\sigma_0 = P/\pi r_0^2$ ) for a bead subject to uniformly distributed compression load (right inset). **c**, Compressive stress ( $\sigma_0$ ) and relative change in green intensity as a function of applied compressive strain for the active difunctional PMMA-4 (red line and red dots) and monofunctional control PMMA-5 (black line and black dots) beads. **d**, Threshold (colour) stress and yield stress are nearly coincident as a function of strain rate, while the threshold strain remains relatively constant.

## METHODS SUMMARY

**Materials.** The spiropyran functionalized PMA was compression moulded into non-standard ‘dog bone’ shaped tensile samples (Supplementary Fig. S1). The PMA material was pressed in an aluminium mould at 54 °C under 100 pounds per square inch (psi) for five minutes, and then cooled to room temperature while holding the pressure at 100 psi. Any thermally induced ring-opening of the mechanophore during processing was reversed by illumination with intense white light. Mechanophore cross-linked PMMA beads were annealed at 90 °C and illuminated with bright visible light for 2.5 h to close the merocyanine to the spiropyran form.

**Mechanical testing.** PMA samples were tested on a screw-driven load frame. The initial gauge length ranged between 2 and 4 mm. For monotonic tension tests, samples were loaded under displacement control at a rate of 1.5  $\text{mm s}^{-1}$  to failure. For fatigue tests, the sample was initially loaded to a maximum extension of twice the gauge length and then unloaded. The maximum extension of subsequent loading cycles was increased by the initial gauge length each time, such that on the second cycle the maximum extension was three times the initial gauge, on the third cycle it was four times the initial gauge length, and so on until the sample failed. Compression tests of spiropyran functionalized PMMA beads were conducted using a custom-built apparatus (Supplementary Fig. S5) at loading rates ranging from 1 to 25  $\mu\text{m s}^{-1}$ . The beads were immersed and tested in oil with a refractive index (1.518) similar to that of PMMA (1.50) to minimize light scattering and improve imaging. A confocal scanning laser microscope (Leica SP2) in fluorescence mode exciting with 543 nm laser light and collecting between 570 and 650 nm was used to obtain fluorescence images.

More detailed descriptions of the materials synthesis and mechanical test protocols are provided in the Supplementary Information.

Received 12 November 2008; accepted 5 March 2009.

- French, A. S. Mechanotransduction. *Annu. Rev. Physiol.* **54**, 135–152 (1992).
- Orr, A. W., Helmke, B. P., Blackman, B. R. & Schwartz, M. A. Mechanisms of mechanotransduction. *Dev. Cell* **10**, 11–20 (2006).

- Gillespie, P. G. & Walker, R. G. Molecular basis of mechanosensory transduction. *Nature* **413**, 194–202 (2001).
- Lele, T. P., Thodeti, C. K. & Ingber, D. E. Force meets chemistry: Analysis of mechanochemical conversion in focal adhesions using fluorescence recovery after photobleaching. *J. Cell. Biochem.* **97**, 1175–1183 (2006).
- Mahadevan, L. & Matsudaira, P. Motility powered by supramolecular springs and ratchets. *Science* **288**, 95–99 (2000).
- Martinac, B. Mechanosensitive ion channels: molecules of mechanotransduction. *J. Cell Sci.* **117**, 2449–2460 (2004).
- Beyer, M. K. & Clausen-Schaumann, H. Mechanochemistry: The mechanical activation of covalent bonds. *Chem. Rev.* **105**, 2921–2948 (2005).
- Löwe, C. & Weder, C. Oligo(*p*-phenylene vinylene) excimers as molecular probes: deformation-induced color changes in photoluminescent polymer blends. *Adv. Mater.* **14**, 1625–1629 (2002).
- Kim, S.-J. & Reneker, D. H. A mechanochromic smart material. *Polym. Bull.* **31**, 367–374 (1993).
- Nallicheri, R. A. & Rubner, M. F. Investigations of the mechanochromic behavior of poly(urethane diacetylene) segmented copolymers. *Macromolecules* **24**, 517–525 (1991).
- Foulger, S. H. *et al.* Mechanochemical response of poly(ethylene glycol) methacrylate hydrogel encapsulated crystalline colloidal arrays. *Langmuir* **17**, 6023–6026 (2001).
- Foulger, S. H. *et al.* Photonic crystal composites with reversible high-frequency stop band shifts. *Adv. Mater.* **15**, 685–689 (2003).
- Comrie, J. E. & Huck, W. T. S. Exploring actuation and mechanotransduction properties of polymer brushes. *Macromol. Rapid Commun.* **29**, 539–546 (2008).
- Azzaroni, O. *et al.* Mechanically induced generation of counterions inside surface-grafted charged macromolecular films: Towards enhanced mechanotransduction in artificial systems. *Angew. Chem. Int. Edn Engl.* **45**, 7440–7443 (2006).
- Hickenboth, C. R. *et al.* Biasing reaction pathways with mechanical force. *Nature* **446**, 423–427 (2007).
- Potisek, S. L. *et al.* Mechanophore-linked addition polymers. *J. Am. Chem. Soc.* **129**, 13808–13809 (2007).
- Tipikin, D. S. Mechanochromism of organic compounds by the example of spiropyran. *Russ. J. Phys. Chem.* **75**, 1720–1722 (2001).
- Minkin, V. I. Photo-, thermo-, solvato-, and electrochromic spiroheterocyclic compounds. *Chem. Rev.* **104**, 2751–2776 (2004).

19. Percec, V. *et al.* Ultrafast synthesis of ultrahigh molar mass polymers by metal-catalyzed living radical polymerization of acrylates, methacrylates, and vinyl chloride mediated by SET at 25 degrees C. *J. Am. Chem. Soc.* **128**, 14156–14165 (2006).
20. Casale, A. & Porter, R. S. *Polymer Stress Reactions* Vol. 1, 8–80 and 96–101 (Academic Press, 1978).
21. Ben-Nun, M. & Martinez, T. J. *Ab initio* quantum molecular dynamics. *Adv. Chem. Phys.* **121**, 439–512 (2002).
22. Sotomayor, M. & Schulten, K. Single-molecule experiments *in vitro* and *in silico*. *Science* **316**, 1144–1148 (2007).
23. Saitta, A. M. & Klein, M. L. First-principles molecular dynamics study of the rupture processes of a bulklike polyethylene knot. *J. Phys. Chem. B* **105**, 6495–6499 (2001).
24. Beyer, M. K. The mechanical strength of a covalent bond calculated by density functional theory. *J. Chem. Phys.* **112**, 7307–7312 (2000).
25. Hiramatsu, Y. & Oka, Y. Determination of the tensile strength of rock by a compression test of an irregular test piece. *Int. J. Rock Mech. Mining Sci.* **3**, 89–99 (1966).
26. Sternberg, E. & Rosenthal, F. The elastic sphere under concentrated loads. *J. Appl. Mech.* **19**, 413–421 (1952).
27. Eyring, E. Viscosity, plasticity, and diffusion as examples of absolute reaction rates. *J. Chem. Phys.* **4**, 283–291 (1936).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We acknowledge the support of the ARO MURI programme (grant number W911NF-07-1-0409) for this research.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to N.R.S. ([n-sottos@illinois.edu](mailto:n-sottos@illinois.edu)).



# Self-assembly of a nanoscale DNA box with a controllable lid

Ebbe S. Andersen<sup>1,2,3</sup>, Mingdong Dong<sup>1,2,4,†</sup>, Morten M. Nielsen<sup>1,2,3</sup>, Kasper Jahn<sup>1,2,3</sup>, Ramesh Subramani<sup>1,2,4</sup>, Wael Mamdouh<sup>1,2,4</sup>, Monika M. Golas<sup>5,8</sup>, Bjoern Sander<sup>6,8</sup>, Holger Stark<sup>8,9</sup>, Cristiano L. P. Oliveira<sup>2,7</sup>, Jan Skov Pedersen<sup>2,7</sup>, Victoria Birkedal<sup>2</sup>, Flemming Besenbacher<sup>1,2,4</sup>, Kurt V. Gothelf<sup>1,2,7</sup> & Jørgen Kjems<sup>1,2,3</sup>

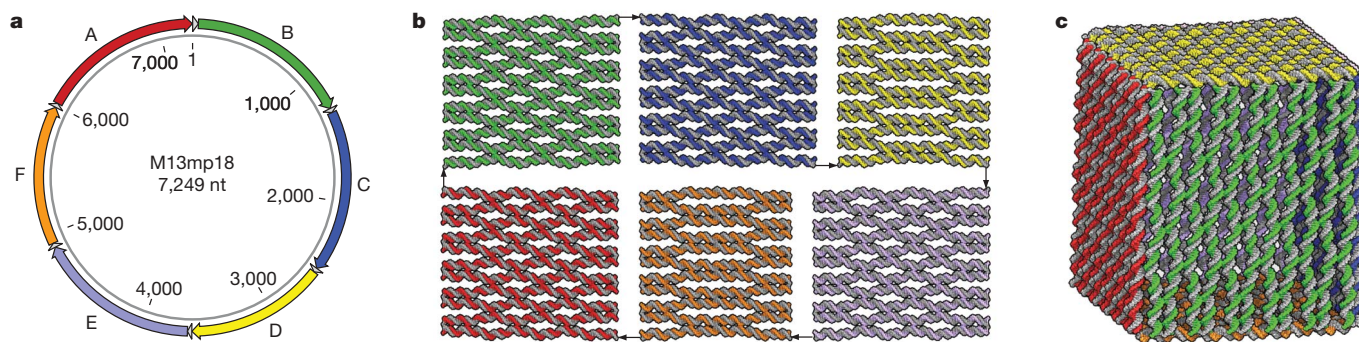
The unique structural motifs and self-recognition properties of DNA can be exploited to generate self-assembling DNA nanostructures of specific shapes using a 'bottom-up' approach<sup>1</sup>. Several assembly strategies have been developed for building complex three-dimensional (3D) DNA nanostructures<sup>2–8</sup>. Recently, the DNA 'origami' method was used to build two-dimensional addressable DNA structures of arbitrary shape<sup>9</sup> that can be used as platforms to arrange nanomaterials with high precision and specificity<sup>9–13</sup>. A long-term goal of this field has been to construct fully addressable 3D DNA nanostructures<sup>14,15</sup>. Here we extend the DNA origami method into three dimensions by creating an addressable DNA box  $42 \times 36 \times 36 \text{ nm}^3$  in size that can be opened in the presence of externally supplied DNA 'keys'. We thoroughly characterize the structure of this DNA box using cryogenic transmission electron microscopy, small-angle X-ray scattering and atomic force microscopy, and use fluorescence resonance energy transfer to optically monitor the opening of the lid. Controlled access to the interior compartment of this DNA nanocontainer could yield several interesting applications, for example as a logic sensor for multiple-sequence signals or for the controlled release of nanocargos.

We designed the DNA box by using a recently developed software package<sup>16</sup> to fold six DNA origami sheets along the circular, single-stranded DNA genome of the M13 bacteriophage (faces indicated with the letters A–F, Fig. 1a). The software built atomic models of the six interconnected sheets (Fig. 1b), which were subsequently

arranged to form a 3D box (Fig. 1c). We then constructed the staple strands that fold the box by bridging the edges, resulting in a 'cuboid' structure of external size  $42 \times 36 \times 36 \text{ nm}^3$  (design details in Supplementary Notes 1–3). This particular design was chosen for several reasons: to use the entire M13 sequence, to ensure a circular folding path through the faces and to introduce faces that have the characteristics of lids with 'hinges' composed of scaffold linkers. Finally, the lid was functionalized with a lock–key system to control its opening (Supplementary Fig. 1).

The designed DNA structure formed by self-assembly after we heat annealed the 220 staple strands onto the single-stranded M13 DNA, resulting in highly homogenous structures migrating as one distinct band in native gel electrophoresis (Supplementary Fig. 2). Initially, a sample without the edge-linking staple strands was imaged in liquid using atomic force microscopy (AFM), revealing that the six DNA sheets formed efficiently and, in most instances, aligned in two parallel rows (Fig. 2a), which is compatible with helical stacking interactions between the edges of the DNA sheets<sup>9</sup> (Supplementary Fig. 3a).

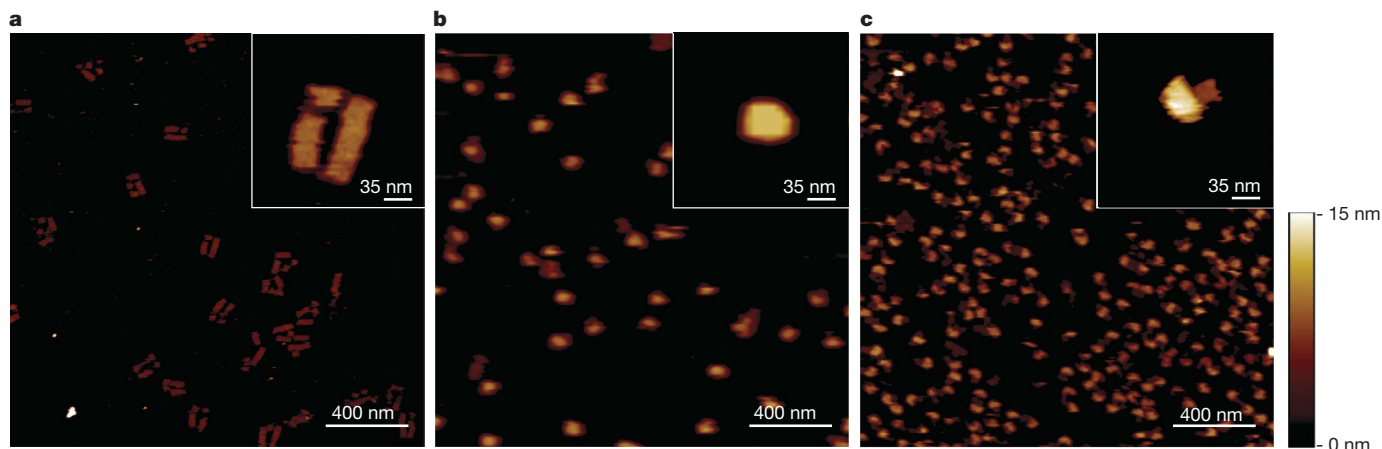
In a second assembly reaction, we used 59 staple strands, connecting the edges, to form the box shape, and subsequent AFM imaging revealed box-like particle structures (Fig. 2b). Analysis of the high-resolution AFM images of individual particles revealed *x* and *y* dimensions that were in good agreement with the shape and dimensions of the designed DNA box. However, the measured heights of these particles varied between 4 and 12 nm (Supplementary Fig. 3b),



**Figure 1 | Design of a DNA origami box.** **a**, Sequence map of the circular, single-stranded DNA genome of the M13 bacteriophage with regions used to fold the six DNA sheets shown as coloured arrows (A–F). Base numbering starts from a 44-nucleotide spacer region between sheets A and B that

contains a stable hairpin structure<sup>9</sup>. Spacers of 33 nucleotides are positioned between each face. **b**, **c**, Molecular models of the six DNA sheets in a flat and cubic higher-order structure, respectively. The six DNA sheets are colour-coded as in **a**.

<sup>1</sup>Danish National Research Foundation: Centre for DNA Nanotechnology, <sup>2</sup>Interdisciplinary Nanoscience Center, <sup>3</sup>Department of Molecular Biology, <sup>4</sup>Department of Physics and Astronomy, <sup>5</sup>The Water and Salt Research Center, Institute of Anatomy, <sup>6</sup>Stereology and EM Research Laboratory, <sup>7</sup>Department of Chemistry, Aarhus University, DK-8000 Aarhus, Denmark. <sup>8</sup>Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, D-37077 Göttingen, Germany. <sup>9</sup>Göttingen Centre for Molecular Biology, Justus-von-Liebig-Weg 11, University of Göttingen, D-37077 Göttingen, Germany. <sup>†</sup>Present address: Rowland Institute at Harvard, Harvard University, 100 Edwin H. Land Boulevard, Cambridge, Massachusetts 02142, USA.



**Figure 2 | AFM imaging of two- and three-dimensional DNA origami structures.** **a**, AFM image of a sample in which the six DNA sheets were folded along the M13 backbone. Inset, magnified view of a preferred arrangement of the six sheets. **b**, AFM image of a sample in which the edges of the DNA sheets were linked to form a box. Inset, magnified view of a box-

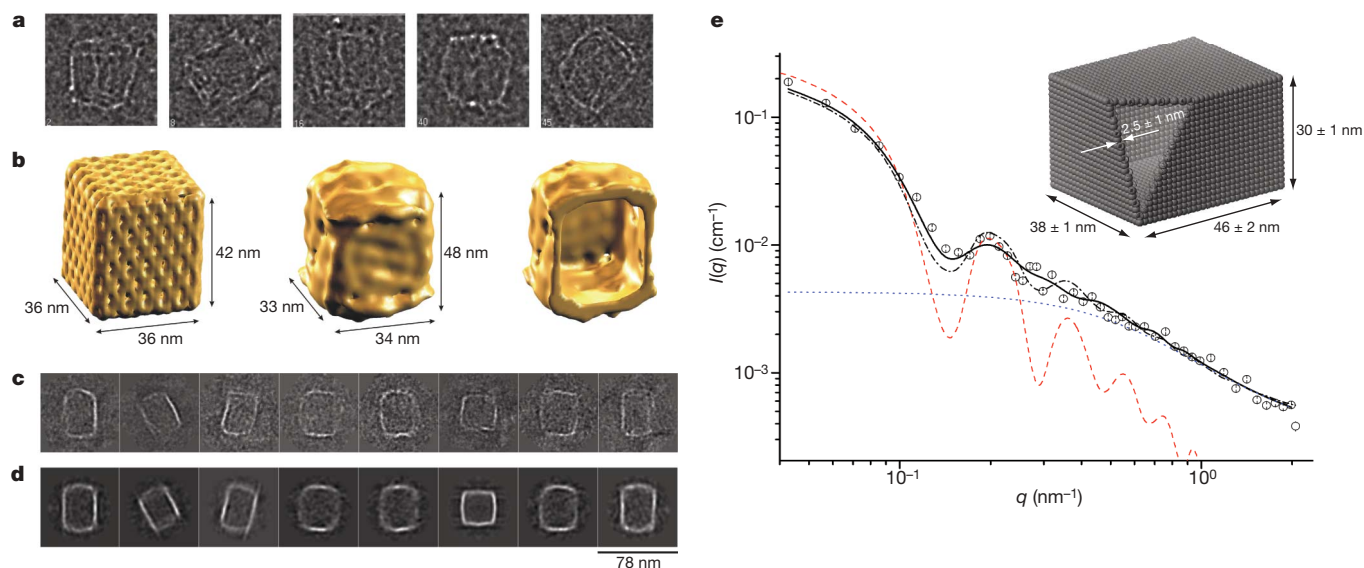
which is most likely caused by distortion of the hollow, soft DNA box by the force exerted by the AFM tip. In support of this argument, we observed that prolonged scanning of individual 3D DNA structures results in their collapse into flat structures one or two DNA sheets in height (Supplementary Fig. 4a).

To further investigate the predictability of the 3D folding, we assembled an open DNA box in which the staple strands closing lid D were omitted. AFM images revealed flat particles two DNA sheets in height with a one-sheet extension of dimensions similar to lid D (Fig. 2c), which probably corresponds to a sideways collapse of the structure (Supplementary Fig. 4b). In conclusion, the AFM images obtained before and after addition of the staple strands connecting the edges of the DNA box strongly indicate the formation of box-shaped structures.

To gain stronger evidence for the formation of the DNA box, we performed cryogenic transmission electron microscopy (cryo-EM) on

like particle. **c**, AFM image of a sample in which one lid of the DNA box was left open. Inset, magnified view of a structure in which the lid is protruding from the body of the box. The colour scale shows the height above the surface in the range 0–15 nm.

a sample of the assembly. Individual DNA boxes showed prominent edges (Fig. 3a) resulting from the increased density of the lateral faces oriented more or less parallel to the incident electron beam. Nearly all of the individual particles (90–95%) seen in the cryo-EM images had an overall box-like shape. A total of 8,987 defocus-corrected box images were subjected to single-particle image processing, and class averages with ~10–30 members on average showed the main structural features of the DNA box (Fig. 3c). For 337 class averages, electron-dense edges could be used to measure the short and long side lengths of the box, yielding dimensions of  $44.6 \pm 3.7$  nm by  $36.8 \pm 3.2$  nm (Supplementary Fig. 5). For 3D reconstruction, class averages were selected using the 3D error as selection criterion (Fig. 3b). The shape of the 3D cryo-EM map of the DNA box was consistent with the class averages and the corresponding two-dimensional re-projections (Fig. 3c, d). Overall, the 3D cryo-EM map revealed a hollow box-like structure with slightly convex bending of the small faces and slightly



**Figure 3 | Characterization of DNA origami box by cryo-EM and small-angle X-ray scattering (SAXS).** **a**, Single-particle cryo-EM images of box-shaped assemblies. **b**, Single-particle reconstruction of the DNA box applying D2 symmetry. Left, theoretical model. Middle, surface representation of the cryo-EM map. Right, cut-open view showing the interior cavity of the cryo-EM map. **c**, **d**, Comparison of the class averages of the DNA boxes (**c**) with the corresponding two-dimensional re-projections of the 3D cryo-EM map (**d**). **e**, Experimental SAXS data (circles) with

corresponding fits from different approaches: red dashed curve, fit using the theoretical atomic coordinates for the box; blue dotted curve, typical Debye background<sup>25</sup> for modelling the excess oligonucleotides; dash-dot curve, fit using the theoretical atomic coordinates for the box with the Debye background added; solid curve, fit using a semi-analytical model for a box with three different side lengths with the Debye background added.  $I(q)$ , SAXS intensity;  $q$ , momentum transfer modulus (Methods). Inset, semi-analytical box model with the estimated side lengths and wall thickness.



concave bending of the large faces (Fig. 3b; see Supplementary Fig. 6 for additional views of the 3D map). We suggest that the convex and concave bending of the faces may reflect the difference in design of the small and large faces. The dimensions of the 3D reconstruction are compatible with 82% of the measured class averages, whereas the theoretical model is compatible with 61% of them (Supplementary Fig. 5). We conclude that the main fraction of the self-assembled structures had a hollow box-like shape that is very similar to the intended 3D design.

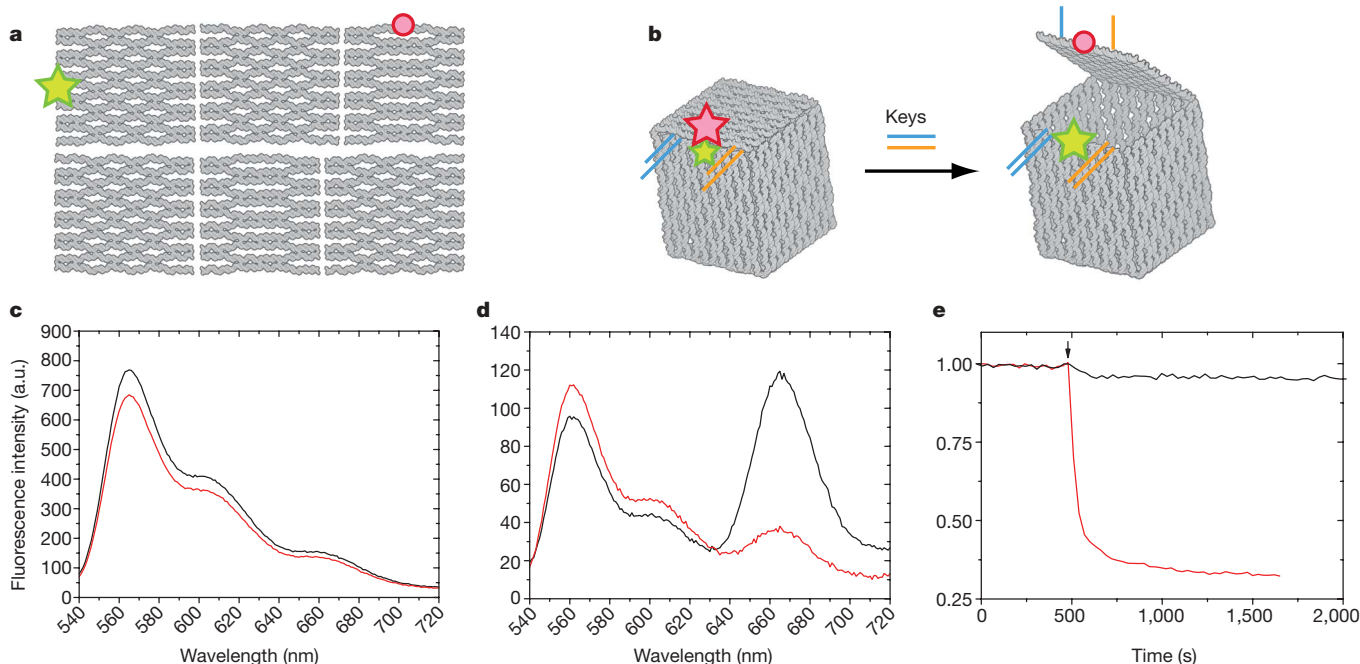
To analyse the native DNA origami box in solution and without a potentially disruptive sample fixation, we used dynamic light scattering and SAXS, which probe a large ensemble and can therefore determine whether or not the box is the major product of the self-assembly reaction. Analysis by dynamic light scattering showed there to be a contribution with a hydrodynamic radius  $R_h = 24 \pm 4$  nm, which matches the calculated value of 25.2 nm from the atomic model, and a contribution with  $R_h = 2.1$  nm, corresponding to non-annealed DNA oligonucleotides (Supplementary Fig. 7). The SAXS analysis yielded a characteristic profile of a well-defined particle and Fourier transformation analysis gave a single-particle self-correlation function characteristic of a hollow structure (data not shown). The scattering data was in good agreement with the atomic model of the DNA box structure (Fig. 3e). We also compared the experimental data with a theoretical model of a box with dimensions  $2a$ ,  $2b$  and  $2c$  and wall thickness  $t$  (Methods and Supplementary Fig. 8). This model can satisfactorily describe the scattering data and gave overall dimensions of  $46 \pm 2$  nm by  $38 \pm 1$  nm by  $30 \pm 1$  nm and a wall thickness of  $2.5 \pm 1$  nm, which is close to the expected size and is consistent with the dimensions observed using cryo-EM. In conclusion, the SAXS data showed that the scattering derives mainly from box-shaped structures, providing further evidence of the successful self-assembly of 3D hollow boxes of well-defined dimensions in solution.

Previous studies have shown that DNA nanostructures can be dynamically manipulated by external DNA sequence signals<sup>7,17,18</sup>. Here we functionalized lid D of the DNA box with a dual lock–key

system composed of DNA duplexes with sticky-end extensions to provide a ‘toehold’<sup>19</sup> for the displacement by externally added ‘key’ oligonucleotides (Supplementary Fig. 1). To detect the opening process of the DNA box lid, we inserted two fluorescent dyes, Cy3 and Cy5, into faces B and D, respectively (Fig. 4a and Supplementary Note 4). Efficient fluorescence resonance energy transfer (FRET) between the two fluorophores corresponds to a closed state in which the dyes are in close proximity (Fig. 4b, left). The addition of key oligonucleotides results in the opening of the lid, resulting in a decrease in the FRET efficiency as the distance between the two dyes increases (Fig. 4b, right).

The functionalized DNA box was assembled, purified and subjected to ensemble FRET measurements. We first studied the six faces without edge links and found emission only from primary excited Cy3, and only a dilution effect was observed upon addition of key oligonucleotides (Fig. 4c). By contrast, emission from the closed box sample had a Cy5 fluorescence peak at a wavelength of 665 nm that is consistent with a FRET signal with an estimated efficiency of  $\sim 0.22$  (Fig. 4d). The addition of both keys led to a decreased Cy5 signal and an increased Cy3 signal (Fig. 4d), and the FRET efficiency decreased by 90%, to  $\sim 0.02$ . The difference in response of the Cy5 and Cy3 signals may be caused by the difference in local environment of the two fluorophores. Cy5 was positioned on the side of a DNA helix, whereas Cy3 was positioned at the end of a DNA helix, where base-stacking interactions<sup>20</sup> may partly quench the fluorophore.

We measured the kinetics of the opening process (Fig. 4e) and found the FRET signal to decrease biexponentially, with an initial (fast) decay time of  $\sim 40$  s upon key addition. The effect was specific, as no significant reduction was observed upon addition of an unrelated oligonucleotide (Fig. 4e). Order-of-addition experiments showed that both keys are required for full decrease of the FRET signal (Supplementary Fig. 9). This indicates that a closed box can be programmed to open in response to at least two external signals (representing an AND gate). The box lid could potentially be designed to close again in the presence of specific signals (representing a NOT gate), and because the DNA box



**Figure 4 | Programmed opening of the box lid.** **a, b**, Illustrations of the unlinked faces of the box (**a**) and the controlled opening of the box lid (**b**). The emission from the Cy5 and Cy3 fluorophores are marked with red and green stars, respectively. Loss of emission from Cy5 is denoted by a red circle and the independent lock–key systems are indicated in blue and orange. **c**, Ensemble FRET measurements of the unlinked faces before (black

curve) and 12 min (red curve) after the addition of keys. **d**, Ensemble FRET measurements of the closed box before (black curve) and 35 min after (red curve) the addition of keys. **e**, Kinetic study of change in emission of Cy5. Black arrow, time of addition of key oligonucleotides (red curve) or an unrelated oligonucleotide (black curve). Initial fluorescence was normalized to one. a.u., arbitrary units.

has two similar lids, D and F, they may be programmed individually (representing an OR gate). Thus, the lids of the DNA box have the potential to be uniquely programmed to respond to complex combinations of oligonucleotide sequences, for example cellular messenger RNAs or micro RNAs.

We have designed, synthesized and thoroughly characterized a megadalton-sized DNA box with a cavity that is large enough to contain, for example, a ribosome or a poliovirus. Furthermore, we have demonstrated dynamic control and programmability of the box lid. In an earlier study, we demonstrated that DNA origami structures can be designed that have tunable flexible properties<sup>16</sup>, and here we further show that dynamic changes can be induced by sensing external signals in the environment. The application of such a 'nanorobotic' device could be to restrict the transport of material in or out of the box in a controlled fashion. Importantly, the opening mechanism of the DNA box operates under native conditions, which allows biologically active components like enzymes to be packaged to provide control of access to their relevant substrates. We note that the DNA box presented here has the potential to both sense and act, for example by combining a diagnostic sensor of complex signals with the controlled release of, or access to, a payload.

## METHODS SUMMARY

**Software.** The software package used for the box design consists of a sequence editor and an extendable algorithm toolbox<sup>16</sup>. We developed a program for creating realistic 3D models, which facilitated the design of the 3D edge-to-edge staple strand crossovers. The software package is distributed as free software under the GNU General Public License version 3 (GPLv3) and documentation and tutorials are available at [www.cdna.dk/origami](http://www.cdna.dk/origami). A specific tutorial for designing the DNA box is available (see *Nature Protocols* doi:10.1038/nprot.2009.75).

**AFM.** We deposited the samples on a mica surface and performed the AFM imaging in a buffer solution. The best AFM images of the DNA box were obtained when applying minimal loading forces and using optimized feedback parameters.

**Cryo-EM.** The sample was adsorbed on carbon film and plunge-frozen in liquid ethane. We made low-dose images under cryogenic conditions using a  $4k \times 4k$  charge-coupled-device detector at 0.245 nm per pixel (ref. 21), and a total of 8,987 single-particle images were selected manually and analysed in the context of IMAGIC-5 (ref. 22).

**SAXS.** We collected data using a high-resolution set-up of a laboratory-based SAXS instrument<sup>23</sup>. The scattering data was analysed on the basis of a theoretical model for a rectangular box using home-written software, and with an atomic-coordinate-based model applying the program CRY SOL<sup>24</sup>.

**FRET.** To detect the opening process, we functionalized the box with Cy3 and Cy5 fluorophores and a lock-key system. Purified samples were investigated using ensemble FRET spectroscopy measurements. The opening of the box was directly monitored in fluorescence kinetic experiments, where DNA key oligonucleotides were added to the closed-box sample.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 9 November 2008; accepted 6 March 2009.

- Seeman, N. C. An overview of structural DNA nanotechnology. *Mol. Biotechnol.* **37**, 246–257 (2007).
- Chen, J. H. & Seeman, N. C. Synthesis from DNA of a molecule with the connectivity of a cube. *Nature* **350**, 631–633 (1991).
- Shih, W. M., Quispe, J. D. & Joyce, G. F. A 1.7-kilobase single-stranded DNA that folds into a nanoscale octahedron. *Nature* **427**, 618–621 (2004).

- Goodman, R. P. et al. Rapid chiral assembly of rigid DNA building blocks for molecular nanofabrication. *Science* **310**, 1661–1665 (2005).
- Douglas, S. M., Chou, J. J. & Shih, W. M. DNA-nanotube-induced alignment of membrane proteins for NMR structure determination. *Proc. Natl Acad. Sci. USA* **104**, 6644–6648 (2007).
- Andersen, F. F. et al. Assembly and structural analysis of a covalently closed nanoscale DNA cage. *Nucleic Acids Res.* **36**, 1113–1119 (2008).
- Yang, H. & Sleiman, H. F. Templated synthesis of highly stable, electroactive, and dynamic metal-DNA branched junctions. *Angew. Chem. Int. Ed.* **47**, 2443–2446 (2008).
- He, Y. et al. Hierarchical self-assembly of DNA into symmetric supramolecular polyhedra. *Nature* **452**, 198–201 (2008).
- Rothemund, P. W. Folding DNA to create nanoscale shapes and patterns. *Nature* **440**, 297–302 (2006).
- Williams, B. A., Lund, K., Liu, Y., Yan, H. & Chaput, J. C. Self-assembled peptide nanoarrays: an approach to studying protein-protein interactions. *Angew. Chem. Int. Ed.* **46**, 3051–3054 (2007).
- Ke, Y., Lindsay, S., Chang, Y., Liu, Y. & Yan, H. Self-assembled water-soluble nucleic acid probe tiles for label-free RNA hybridization assays. *Science* **319**, 180–183 (2008).
- Rinker, S., Ke, Y., Liu, Y., Chhabra, R. & Yan, H. Self-assembled DNA nanostructures for distance-dependent multivalent ligand-protein binding. *Nature Nanotechnol.* **3**, 418–422 (2008).
- Sharma, J. et al. Toward reliable gold nanoparticle patterning on self-assembled DNA nanoscaffold. *J. Am. Chem. Soc.* **130**, 7820–7821 (2008).
- Seeman, N. C. Nucleic acid junctions and lattices. *J. Theor. Biol.* **99**, 237–247 (1982).
- Simmel, F. C. Three-dimensional nanoconstruction with DNA. *Angew. Chem. Int. Ed.* **47**, 5884–5887 (2008).
- Andersen, E. S. et al. DNA origami design of dolphin-shaped structures with flexible tails. *ACS Nano* **2**, 1213–1218 (2008).
- Ding, B. & Seeman, N. C. Operation of a DNA robot arm inserted into a 2D DNA crystalline substrate. *Science* **314**, 1583–1585 (2006).
- Goodman, R. P. et al. Reconfigurable, braced, three-dimensional DNA nanostructures. *Nature Nanotechnol.* **3**, 93–96 (2008).
- Yurke, B., Turberfield, A. J., Mills, A. P., Simmel, F. C. & Neumann, J. L. A DNA-fuelled molecular machine made of DNA. *Nature* **406**, 605–608 (2000).
- Iqbal, A. et al. Orientation dependence in fluorescent energy transfer between Cy3 and Cy5 terminally attached to double-stranded nucleic acids. *Proc. Natl Acad. Sci. USA* **105**, 11176–11181 (2008).
- Sander, B., Golas, M. M. & Stark, H. Advantages of CCD detectors for de novo three-dimensional structure determination in single-particle electron microscopy. *J. Struct. Biol.* **151**, 92–105 (2005).
- van Heel, M., Harauz, G., Orlova, E. V., Schmidt, R. & Schatz, M. A new generation of the IMAGIC image processing system. *J. Struct. Biol.* **116**, 17–24 (1996).
- Pedersen, J. S. A flux- and background-optimized version of the NanoSTAR small-angle X-ray scattering camera for solution scattering. *J. Appl. Crystallogr.* **37**, 369–380 (2004).
- Svergun, D., Barberato, C. & Koch, M. H. J. CRY SOL - A program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.* **28**, 768–773 (1995).
- Debye, P. Molecular-weight determination by light scattering. *J. Phys. Colloid Chem.* **51**, 18–32 (1947).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank R. Rosendahl Hansen and J. Kristensen for technical assistance. This work was supported by grants from the Danish National Research Foundation to the Centre for DNA Nanotechnology and the Danish Research Agency through support to the Interdisciplinary Nanoscience Center, by the Federal Ministry of Education and Research, Germany (0311899), and by the Sixth Framework Program of the European Union through the Integrated Project '3D Repertoire' (H.S.).

**Author Information** The 3D map has been deposited in the EM Data Bank under the accession code EMD-1612. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to K.V.G. (kvg@chem.au.dk) or J.K. (jk@mb.au.dk).



## METHODS

**Sample preparation.** We prepared the m13mp18 DNA as described previously<sup>16</sup>. The assembly reactions were performed in Tris-acetate-EDTA buffer with 12.5 mM MgAc (TAEM), 1.6 mM M13 and fivefold excess of each oligonucleotide. The samples were heated to 95 °C and cooled to 20 °C in steps of 0.1 °C every 6 s.

**AFM.** The sample (5 µl) was deposited on a freshly cleaved mica surface (Ted Pella) and left to adsorb for 5 min. Then 200 µl TAEM buffer was added to the liquid cell and the sample was scanned in tapping-mode using either an Agilent AFM series 5500 (Agilent Technologies) or a MultiMode SPM with a Nanoscope IV controller (Veeco Instruments) under ambient conditions. We used oxide-sharpened triangular cantilevers (NP-S, Veeco Instruments) with a spring constant of 0.4 N m<sup>-1</sup> and a normal tip radius of 10 nm. After engagement, the tapping-amplitude set point was typically 0.5 V and the scan rates ranged from 1 to 2 Hz. During the liquid AFM imaging, the highest resolution was in most cases obtained with minimal loading forces applied and using optimized feedback parameters. We obtained several AFM images, all 512 × 512 pixels, from separate locations across the mica surfaces to ensure reproducibility of the results. All the AFM images were analysed using Scanning Probe Image Processor (SPIP) software (Image Metrology ApS, version 4.7.4).

**Sample preparation for cryo-EM.** For cryo-EM, we supplemented the DNA box sample with 1 mM spermidine. For adsorption of the particles, a thin carbon film was used that was prepared by evaporating carbon on a piece of freshly cleaved mica. About 25 µl of the sample were filled into a well in an ice-cooled block, and particles were adsorbed on a carbon film for between 30 min and 12 h. Subsequently, the floating carbon film was picked out from the particle solution by placing a copper EM grid covered with a perforated carbon film on top of the floating carbon film. After adding 5 µl of the particle solution on one side of the EM grid, we mounted the grid in a home-made freeze-plunging device, blotted and immediately plunge-froze it in liquid ethane. The EM grids were stored in liquid nitrogen until image acquisition.

**Electron cryomicroscopy.** Specimen grids were transferred to the electron cryomicroscope (CM200 FEG electron microscope, Philips/FEI) under cryogenic conditions using a cryoholder (Gatan). The electron microscope was operated at an acceleration voltage of 160 kV in low-dose mode. Images were taken on a 4k × 4k CCD (charge-coupled-device) camera (TemCam-F415, TVIPS) with twofold binning of the pixels and ×122,000 magnification (ref. 21). A total of 4,160 CCD camera exposures were collected.

**Image processing.** Image processing was performed in the framework of the IMAGIC-5 software package<sup>22</sup>. We manually selected 8,987 single-particle images of the DNA boxes, and corrected them for the contrast transfer function as described previously<sup>26</sup>. For image processing, the images were coarsened to a pixel size of 4.9 Å at the specimen level. Images were subjected to three iterative rounds of multireference alignment using resampling to polar coordinates<sup>27</sup> and multivariate statistical analysis. Aligned images were classified using hierarchical ascendant classification with moving-element refinement, resulting in class averages with ~10–30 class members on average. The 3D reconstruction was iteratively refined in three rounds of projection matching (pixel size, 9.8 Å), starting with 337 class averages that showed well-defined structural features. The 3D error was used to select the best class averages for 3D calculation. Thereby, the number of selected class averages was reduced to 250 and finally to 36.

**Determination of dimensions.** For a graphical and statistical description of the size distribution of the DNA boxes (given as average plus/minus standard deviation), the long and short side lengths were measured for 337 experimental class averages that showed well-defined edges. These values were colour-coded according to their abundance and plotted graphically. The theoretical box model and the experimental 3D structure were re-projected into 1,654 two-dimensional, equally spaced re-projections that were measured to define the respective size distributions.

**Dynamic light scattering.** The experiments were performed at a scattering angle of 90° on an ALV instrument with a CGS-8F goniometer system equipped with an ALV-6010/EPP multi-tau digital correlator and a helium–neon diode laser (JDS Uniphase). ALV software was used to derive the hydrodynamic radius using cumulant analysis and inverse Laplace transforms. The samples were the same as used in the SAXS experiments, but measured at 23 °C.

**SAXS.** The experiments were performed in the laboratory-based SAXS instrument at the Department of Chemistry, Aarhus University<sup>23</sup>. The SAXS intensity was displayed as a function of the momentum transfer modulus  $q$  ( $q = 4\pi\sin(\theta)/\lambda$ ,

where  $\lambda$  is the radiation wavelength and  $2\theta$  is the scattering angle). Pure water was used as primary standard for absolute scale normalization. The sample concentration was very low (~0.1 mg ml<sup>-1</sup>) and the data was collected in the high-resolution set-up covering  $0.004 \text{ Å}^{-1} < q < 0.21 \text{ Å}^{-1}$ . The samples were injected into carefully rinsed quartz capillaries and placed in a thermostated block (4 °C) in the vacuum chamber of the instrument.

The modelling of the scattering data was performed using the expression

$$I(q) = S_{C_1} P_{\text{BOX}}(q) + S_{C_2} P_{\text{DEB}}(q) + \text{Back}$$

where  $S_{C_i}$  are the scale factors for the scattering components and Back is a constant background that might correct imperfect data reduction or the contribution from electron density fluctuations in the sample.  $P_{\text{BOX}}(q)$  is the scattering intensity of a rectangular box with sides  $2a$ ,  $2b$  and  $2c$  and wall thickness  $t$ , given by

$$P_{\text{BOX}}(q) = \frac{2}{\pi} \int_0^{\pi/2} \int_0^{\pi/2} (A_{\text{REC}}(q, a, b, c) - A_{\text{REC}}(q, a-t, b-t, c-t))^2 \sin \alpha \, d\alpha \, d\beta$$

where  $\alpha$  and  $\beta$  are polar angles of integration and  $A_{\text{REC}}(q, a, b, c)$  is the scattering amplitude of a rectangular box with sides  $2a$ ,  $2b$  and  $2c$  (Supplementary Fig. 8), given by

$$A_{\text{REC}} = 8abc \frac{\sin(qa \sin \alpha \sin \beta)}{qa \sin \alpha \sin \beta} \frac{\sin(qb \sin \alpha \cos \beta)}{qb \sin \alpha \cos \beta} \frac{\sin(qc \cos \alpha)}{qc \cos \alpha}$$

$P_{\text{DEB}}(q)$  is the scattering intensity of a Gaussian chain with a radius of gyration  $R_g$  (ref. 25):

$$P_{\text{DEB}}(q) = \frac{2(\exp(-q^2 R_g^2) - 1 + q^2 R_g^2)}{(q^2 R_g^2)^2}$$

Because the sample had free oligonucleotides in the solution (also detected by dynamic light scattering measurements), it was necessary to add this intensity to take into account their contribution to the scattering.

The measured intensity is smeared by instrumental effects. To account for this, we use a resolution function<sup>28</sup>. The observed experimental intensity,  $\tilde{I}(\langle q \rangle)$ , is thus fitted by

$$\tilde{I}(\langle q \rangle) = \int R(\langle q \rangle, q) I(q) \, dq \quad (1)$$

where  $R(\langle q \rangle, q)$  is the resolution function that describes the distribution of  $q$  values probed for the scattering vector setting,  $\langle q \rangle$ . The parameters entering the resolution function are estimated from the width of the direct beam at the detector. Equation (1) is used to fit the experimental data using a nonlinear least-squares procedure<sup>29</sup>.

**FRET.** Samples were purified on HPLC (Agilent Technologies 1200 series) using a BioSep-2000 column (Phenomenex) with an elution buffer containing 20 mM Tris-HCl, pH 8.0, 100 mM KCl and 10 mM MgCl<sub>2</sub>. Ensemble FRET experiments were carried out in a JY-Horiba Fluoromax-3 spectrofluorimeter using a 60-µl quartz cuvette (Hellma UK). Aliquots of 0.2 µl of 200 µM key or an unrelated oligonucleotide were directly added to the box sample and mixed by pipetting. Fluorescence spectra for each sample were recorded at 25 °C with both Cy3 excitation at 530 nm and Cy5 excitation at 600 nm, and corrected for background fluorescence and instrument response. The kinetic studies were performed with a 10-nm slit centred at 665 nm, and recorded each 30 s with high shutter speeds to minimize bleaching. The FRET efficiency was estimated by the RatioA method<sup>30</sup>.

26. Sander, B., Golas, M. M. & Stark, H. Automatic CTF correction for single particles based upon multivariate statistical analysis of individual power spectra. *J. Struct. Biol.* **142**, 392–401 (2003).

27. Sander, B., Golas, M. M. & Stark, H. Corrim-based alignment for improved speed in single-particle image processing. *J. Struct. Biol.* **143**, 219–228 (2003).

28. Pedersen, J. S., Posselt, D. & Mortensen, K. Analytical treatment of the resolution function for small-angle scattering. *J. Appl. Crystallogr.* **23**, 321–333 (1990).

29. Pedersen, J. S. Analysis of small-angle scattering data from colloids and polymer solutions: modeling and least-squares fitting. *Adv. Colloid Interface Sci.* **70**, 171–210 (1997).

30. Clegg, R. M. *et al.* Fluorescence resonance energy transfer analysis of the structure of the four-way DNA junction. *Methods Enzymol.* **211**, 353–388 (1992).

# Upper-mantle volatile chemistry at Oldoinyo Lengai volcano and the origin of carbonatites

T. P. Fischer<sup>1</sup>, P. Burnard<sup>2</sup>, B. Marty<sup>2</sup>, D. R. Hilton<sup>3</sup>, E. Füri<sup>3</sup>, F. Palhol<sup>2</sup>, Z. D. Sharp<sup>1</sup> & F. Mangasini<sup>4</sup>

Carbonatite lavas are highly unusual in that they contain almost no SiO<sub>2</sub> and are >50 per cent carbonate minerals. Although carbonatite magmatism has occurred throughout Earth's history, Oldoinyo Lengai, in Tanzania, is the only currently active volcano producing these exotic rocks<sup>1</sup>. Here we show that volcanic gases captured during an eruptive episode at Oldoinyo Lengai are indistinguishable from those emitted along mid-ocean ridges, despite the fact that Oldoinyo Lengai carbonatites occur in a setting far removed from oceanic spreading centres. In contrast to lithophile trace elements, which are highly fractionated by the immiscible phase separation that produces these carbonatites, volatiles (CO<sub>2</sub>, He, N<sub>2</sub> and Ar) are little affected by this process. Our results demonstrate that a globally homogenous reservoir exists in the upper mantle and supplies volatiles to both mid-ocean ridges and continental rifts. This argues against an unusually C-rich mantle being responsible for the genesis of Na-rich carbonatite and its nephelinite source magma at Oldoinyo Lengai. Rather, these carbonatites are formed in the shallow crust by immiscibility from silicate magmas (nephelinite), and are stable under eruption conditions as a result of their high Na contents.

Oldoinyo Lengai is located on the eastern branch of the East African Rift Valley, away from any influence of a deep-mantle plume<sup>2</sup> or subduction zone. The composite cone consists mainly of nephelinitic and phonolitic pyroclasts and lavas<sup>1</sup>. Although currently erupting more silicic ashes, Oldoinyo Lengai erupted large volumes of rare Na-rich (natro-) carbonatite lavas until August 2007. Such lavas are enriched in incompatible trace elements by up to four orders of magnitude relative to primitive mantle, and show extremely high ( $\geq 10^3$ ) enrichments of light relative to heavy rare-earth elements<sup>3</sup>. The radiogenic isotope compositions of carbonatites, including those emitted by Oldoinyo Lengai, are typical of those that characterize mantle-derived magmas, leaving little doubt that these peculiar magmas are differentiated products of partial melts of the underlying mantle source<sup>4</sup>. Petrologic studies show that natro-carbonatite magma is formed by low-pressure differentiation in the shallow crust from mantle-derived nephelinites<sup>5</sup>. The CO<sub>2</sub>-rich nature (~30 wt%; ref. 1) of carbonatite lavas raises the question of whether mantle-derived melts (for example nephelinites), which can form carbonatite liquids by differentiation, are formed from a typical mantle assemblage under specific conditions (for example extremely low partial-melt fractions<sup>6</sup>) or instead require unusually high C contents in the mantle source (see, for example, ref. 7). Recent experimental results have suggested that nephelinitic magmas, which were probably the Oldoinyo Lengai carbonatite parental liquids, require mantle C contents of ~1,000 p.p.m. at 1–5% partial melting<sup>8</sup>. Logically, a C-enriched mantle source seems necessary to produce liquids that can differentiate into carbonatite magmas.

We used exceptionally uncontaminated volcanic gases captured at Oldoinyo Lengai at the onset of a new eruptive cycle to address the

issue of source C content. We note that the rate at which CO<sub>2</sub> is emitted at Oldoinyo Lengai ( $\sim 5.5 \times 10^{10}$  mol yr<sup>-1</sup>, measured in ref. 9 and by ourselves), is two orders of magnitude greater than the time-averaged solid C (carbonatite) flux ( $5 \times 10^8$  mol yr<sup>-1</sup>; ref. 9). Therefore, the amount of C that forms carbonatite is a minor fraction only (~1%), in comparison with the total C released by the volcano. In this way, the volatile compositions measured in our gas samples are related to the degassing of the whole volcano, and not the volatile phase in equilibrium with carbonatite liquid at low pressure.

The gas samples were collected in July 2005, in the active crater of Oldoinyo Lengai, at the beginning of a new eruptive sequence that resulted in the emission of low-viscosity and low-temperature (550 °C) carbonatite lava flows. Sample collection followed standard protocols<sup>10</sup>. Gas chemistry and N isotope composition were measured by gas chromatography and isotope-ratio mass spectrometry at the University of New Mexico<sup>11</sup>, and noble gases and C isotopes by mass spectrometry at the Scripps Institution of Oceanography<sup>12</sup>. All results are reported in Table 1. To determine the source characteristics of the gases, we used isotopic and elemental ratios, that is, N<sub>2</sub>/Ar, <sup>3</sup>He/<sup>4</sup>He, <sup>40</sup>Ar/<sup>36</sup>Ar and N<sub>2</sub>/<sup>36</sup>Ar, which are diagnostic of magma sources but unaffected by degassing fractionation. The effects of degassing were evaluated from gas abundance ratios (for example He/Ar and CO<sub>2</sub>/Ar), which change upon degassing owing to solubility differences in melts.

There is overwhelming evidence to indicate that Oldoinyo Lengai gas compositions strongly resemble upper-mantle volatiles. First, the relative abundances of N<sub>2</sub>, He and Ar (Fig. 1) overlap with volatiles emitted at mid-ocean ridges and continental rifts. Second, the <sup>3</sup>He/<sup>4</sup>He ratios of Oldoinyo Lengai samples define a narrow range, from 6.70R<sub>A</sub> to 6.86R<sub>A</sub>. They fall at the lower end of the MORB range and are indistinguishable from the subcontinental lithospheric mantle<sup>13</sup>; indeed, there is considerable overlap with well-documented examples of MORB with <sup>3</sup>He/<sup>4</sup>He ratios below 6.5R<sub>A</sub> in slow- to medium-spreading ridges (the southeast and southwest Indian ridges, the south Mid-Atlantic Ridge and the Gakkel Ridge<sup>13</sup>).

Third, the isotopic composition of nitrogen shows  $\delta^{15}\text{N}$  values from -4.0 to -5.1‰ ( $\delta^{15}\text{N}_{\text{air}} = 0.0\text{‰}$ ), similar to MORB and to continental-rift gas values ( $\delta^{15}\text{N}$ ,  $-5 \pm 3\text{‰}$ ; refs 14, 15). Any contribution from N-rich subducted sediments to the Oldoinyo Lengai volatile budget would shift  $\delta^{15}\text{N}$  values towards  $\sim +7\text{‰}$  (refs 11, 16). Fourth, plots of N<sub>2</sub>/<sup>36</sup>Ar against <sup>40</sup>Ar/<sup>36</sup>Ar for the gas samples, along with data for MORB<sup>14</sup>, lie precisely on the mixing line between ASW and the upper mantle (Fig. 2). Although the <sup>40</sup>Ar/<sup>36</sup>Ar ratios of the gases (950) are significantly lower than the upper-mantle end-member (~35,000; ref. 17), they are still among the highest such ratios of any volcanic gases measured. Fifth, the  $\delta^{13}\text{C}$  values (-2.36 to -4.01‰) are within the range of MORB<sup>14</sup> and subcontinental mantle gases<sup>15</sup>.

<sup>1</sup>Department of Earth and Planetary Sciences, MSC03 2040, 1 University of New Mexico, New Mexico 87131-0001, USA. <sup>2</sup>Centre de Recherches Pétrographiques et Géochimiques, Nancy-Université, CNRS-INSU, BP 20, 54501 Vandœuvre Lès Nancy Cedex, France. <sup>3</sup>Geosciences Research Division, Scripps Institution of Oceanography, La Jolla, California 92093-0244, USA. <sup>4</sup>Department of Mining and Mineral Processing Engineering, University of Dar Es Salaam, PO Box 35131, Dar Es Salaam, Tanzania.



**Table 1 | Composition of gas emitted from Oldoinyo Lengai crater, July 2005**

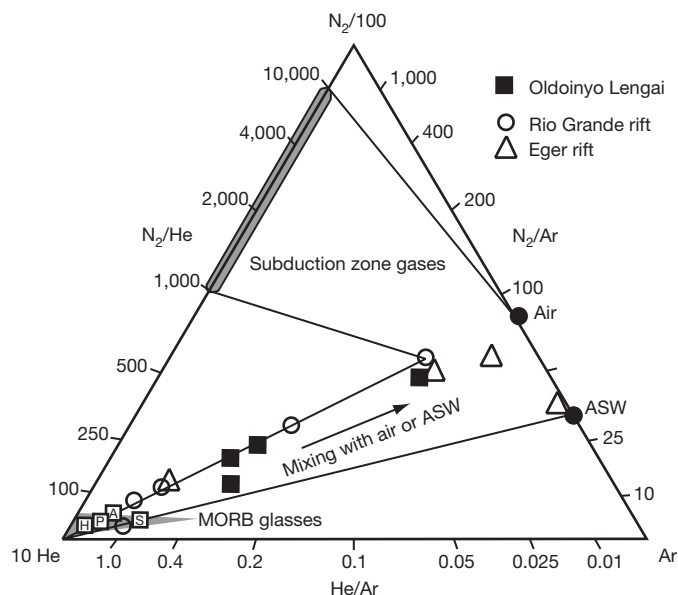
|                               | Sample            |                   |                  |                   |         |               |
|-------------------------------|-------------------|-------------------|------------------|-------------------|---------|---------------|
|                               | OLD-1             | OLD-2             | OLD-3            | OLD-4             | ASW     | MORB          |
| $T$ (°C)                      | 123               | 120               | 168              | 168               | —       | —             |
| $N_2/Ar$                      | 50.4              | 81.9              | 98.1             | 82.6              | 50      | $152 \pm 58$  |
| $N_2/He$                      | 177.2             | 261               | 376.2            | 1,558.3           | 150,000 | $48 \pm 6$    |
| $He/Ar$                       | 0.28              | 0.31              | 0.26             | 0.05              | 0.0005  | 2             |
| $^3He/^4He$ ( $R/R_A$ )       | 6.73              | 6.86              | 6.68             | 6.7               | 1       | $8 \pm 1$     |
| $He/Ne$ (air)                 | 4,600             | 6,530             | 2,020            | 2,370             | 0.78    | 235,000       |
| $^3He/^4He$ ( $R_C/R_A$ )     | $6.73 \pm 0.06$   | $6.86 \pm 0.06$   | $6.68 \pm 0.06$  | $6.70 \pm 0.06$   | —       | $8 \pm 2$     |
| $^{40}Ar/^{36}Ar$             | $290.81 \pm 1.67$ | $947.83 \pm 6.23$ | —                | $317.05 \pm 1.01$ | 295.5   | $\sim 35,000$ |
| $^{20}Ne/^{36}Ar$             | 0.057             | 0.142             | —                | 0.022             | 0.16    | 0.013         |
| $CO_2/^3He$ ( $\times 10^9$ ) | $3.82 \pm 0.08$   | $4.13 \pm 0.08$   | $3.59 \pm 0.08$  | $3.80 \pm 0.08$   | —       | $\sim 1-6$    |
| $N_2/^36Ar$ ( $\times 10^4$ ) | 1.47              | 7.76              | —                | 2.58              | 1.03    | 280           |
| $^4He/^{40}Ar^*$              | —                 | 0.45              | —                | 0.76              | —       | 2             |
| $CO_2/N_2$                    | 203               | 152               | 89               | 23                | —       | 85–190        |
| $CO_2/^{40}Ar^*$              | —                 | 17,886            | —                | 27,166            | —       | 33,600        |
| $\delta^{15}N$ (‰, air)       | $-4.0 \pm 0.2$    | $-4.1 \pm 0.5$    | $-5.1 \pm 0.3$   | —                 | 0       | $-5 \pm 3$    |
| $\delta^{13}C$ (‰, PDB)       | $-2.36 \pm 0.01$  | $-2.40 \pm 0.01$  | $-2.37 \pm 0.01$ | $-4.01 \pm 0.01$  | -8      | $-4.5 \pm 2$  |

Elemental ratios are reported on a molar basis. Errors in the  $N_2/Ar$ ,  $N_2/He$ ,  $He/Ar$  and  $CO_2/N_2$  ratios are 5% (ref. 11). Errors in isotopic values are at the  $1\sigma$  level. The  $He/Ne$  ratios are given relative to the  $He/Ne$  ratio of air. The  $He/Ne$  of the mid-ocean-ridge basalt (MORB) endmember was calculated from the MORB  $^3He/^{22}Ne$  (10.2),  $^3He/^{40}Ar$  ( $8R_A$ ) and  $^{20}Ne/^{22}Ne$  (12.5) ratios; the  $^3He/^{40}Ar$  ratio is reported in  $R_A$  notation, where  $R_A$  is the  $^3He/^{40}Ar$  ratio of air ( $1.4 \times 10^{-6}$ ). Using measured  $He/Ne$  ratios, the measured  $^3He/^{40}Ar$  ( $R$ ) value was corrected for air contamination (details in ref. 12). The  $^{20}Ne/^{36}Ar$  ratio was calculated from the  $He/Ne$ ,  $He/Ar$  and  $^{40}Ar/^{36}Ar$  ratios. The  $^{20}Ne/^{36}Ar$  ratio of air-saturated water (ASW) was measured at 20 °C (ref. 27). The  $^{20}Ne/^{36}Ar$  ratio of the MORB endmember was calculated using the  $^{22}Ne/^{36}Ar$  ratio of 'popping rock'<sup>17</sup> and the  $^{20}Ne/^{22}Ne$  ratio of mantle well gases<sup>25</sup>. The  $CO_2/^3He$  ratio was calculated from the  $CO_2/^3He$ ,  $^3He/^{40}Ar$  and  $^4He/^{40}Ar^*$  ratios.  $^{40}Ar^*$  is  $^{40}Ar$  corrected for atmospheric  $Ar$ . MORB endmember compositions are from refs 13, 14, 16, 28, 29 and 30.  $\delta^{15}N$  (‰) =  $[(^{15}N/^{14}N)_{sample}/(^{15}N/^{14}N)_{standard} - 1] \times 1,000$ ; the standard for N is air.  $\delta^{13}C$  (‰) =  $[(^{13}C/^{12}C)_{sample}/(^{13}C/^{12}C)_{standard} - 1] \times 1,000$ ; the standard for C is Pee Dee Belemnite (PDB).

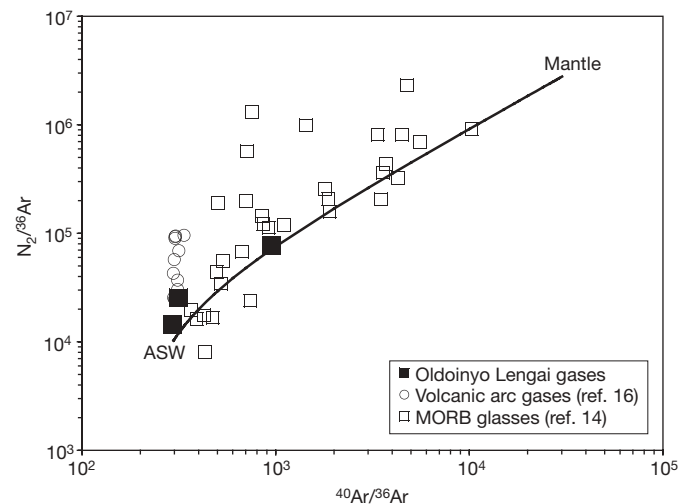
Last, and most surprisingly,  $CO_2/^3He$  ratios of  $(3.6-4.1) \times 10^9$  are within the range of MORB glass values  $((1-6) \times 10^9$ ; ref. 14) and are slightly lower than ratios of subcontinental mantle gases<sup>15,18</sup>. The  $CO_2/^3He$  ratios of Oldoinyo Lengai gases are much lower than values characteristic of gases from arc volcanoes  $((11.0 \pm 3.3) \times 10^9$ ; ref. 19), which are the result of contributions of C-rich components in fluid released from the subducting slab to the mantle wedge.

Elemental noble gas ratios also support this conclusion. The Oldoinyo Lengai  $^4He/^{40}Ar^*$  ratios (0.45–0.76; Fig. 3) are close to the mantle value ( $^4He/^{40}Ar^*$  upper mantle = 2; ref. 17). This observation demonstrates that there was only very minor volatile fractionation during magmatic degassing at Oldoinyo Lengai, because He is  $\sim 10$  times more soluble than Ar in both silicate and carbonatite melts<sup>20,21</sup> and degassing of a magma would result in large increases in the  $He/Ar$  ratio of the residual liquid. Therefore, it is reasonable to assume that the

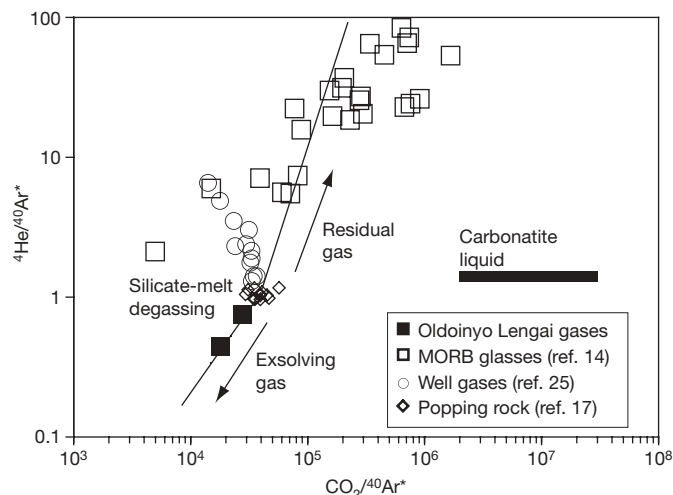
measured volatile compositions are representative of the mantle source. We note that although  $^4He/^{40}Ar^*$  ratios appear unfractionated, the  $CO_2/^3He$  and  $CO_2/^{40}Ar^*$  compositions of the Oldoinyo Lengai gases are not consistent with compositions expected from unfractionated carbonatite liquids. The  $CO_2/^3He$  and  $CO_2/^{40}Ar^*$  ratios of an initial carbonatite liquid (that is, one in equilibrium with an undegassed mantle melt) can be constrained by mass balance to  $\sim 1 \times 10^6$  and  $\sim 2 \times 10^6$ , respectively, which are two orders of magnitude greater than those measured in the gases (Supplementary Information). Thus, enormous degassing fractionation of a carbonatite source would be required to produce the  $CO_2/^3He$  and  $CO_2/^{40}Ar^*$  ratios observed in Oldoinyo Lengai gases, and this is not reflected in the  $^4He/^{40}Ar^*$  ratio. We conclude that there is no degassing scenario from initial carbonatite compositions that could reproduce the Oldoinyo Lengai  $^4He-^{40}Ar^*-CO_2$  compositions. The gases we sampled at Oldoinyo Lengai were derived from degassing of the silicate portion of the magmatic system, as demonstrated by the fact that their compositions plot precisely on the trend predicted by open-system degassing of silicate melts (Fig. 3). This is consistent with the observation that the carbonatite lavas represent only 1% of the Oldoinyo Lengai C budget.



**Figure 1 |  $N_2$ -He-Ar triangular plot.** Oldoinyo Lengai gas compositions and mantle endmembers as defined by MORB glasses<sup>14</sup>, the Rio Grande rift<sup>31</sup>, the Eger rift (Europe)<sup>15</sup>, popping rock (P), the Sierra Negra (S) and Alcedo (A) volcanoes (Galapagos)<sup>32</sup> and Halemaumau crater (H; Kilauea volcano, Hawaii) (T.P.F., unpublished observations). All samples are mixtures between the mantle endmember composition and air or ASW.



**Figure 2 |  $N_2/^{36}Ar$  and  $^{40}Ar/^{36}Ar$  ratio diagram.** Oldoinyo Lengai gases, MORB glasses<sup>14</sup> and arc-gas emissions<sup>16</sup>. All Oldoinyo Lengai samples fall on a mixing trend between ASW and a mantle endmember with  $N_2/^{36}Ar = 280 \times 10^4$  and  $^{40}Ar/^{36}Ar \approx 35,000$  (Table 1).



**Figure 3** |  $^4\text{He}/^{40}\text{Ar}^*$  and  $\text{CO}_2/^{40}\text{Ar}^*$  ratios of Oldoinyo Lengai gases. New Mexico well-gas data<sup>25</sup> and compositions of 21ID43 popping rock are considered representative of the volatile compositions of primary mantle melts<sup>17</sup>. The volatile composition of a carbonatite liquid in equilibrium with such a primary melt is calculated using a mass-balance approach (Supplementary Information). Rayleigh-type degassing will fractionate the exsolved volatiles from the initial volatile compositions (defined by popping rock for silicate magmas and the calculated carbonatite pole for carbonatitic magmas), with the exsolving gases moving towards lower  $^4\text{He}/^{40}\text{Ar}^*$  and  $\text{CO}_2/^{40}\text{Ar}^*$  values owing to higher solubility of He (ref. 21) and  $\text{CO}_2$  (ref. 33) relative to Ar (ref. 21) in the melt. MORB glasses<sup>14</sup> have compositions that represent the residual mantle gas after degassing. The models show that volatiles sampled at Oldoinyo Lengai could not have been exsolved from a carbonatite magma and instead coincide with the compositions expected from a degassing silicate magma.

Using our  $\text{CO}_2/{}^3\text{He}$  ratios, we can constrain the concentration of C in the Oldoinyo Lengai mantle source to be no more than that of the highest C concentrations in the MORB source. The C content of the upper mantle is well constrained by three independent approaches<sup>19,22,23</sup> to fall between 50 and 220 p.p.m. The highest C concentration estimates in this range are found in MORB with  $\text{CO}_2/{}^3\text{He}$  ratios identical to those of Oldoinyo Lengai (Supplementary Information); therefore, the mantle source that is generating Oldoinyo Lengai carbonatite, and its parental nephelinitic magma, contains an absolute maximum of 300 p.p.m. C (including uncertainties in the mantle C estimates). Our results argue against a C-rich mantle being the source of the unusual volcanism at Oldoinyo Lengai, including the Na-rich carbonatites.

Although the Earth's mantle contains significant amounts of C (ref. 19), all present-day volcanoes except Oldoinyo Lengai emit C exclusively as  $\text{CO}_2$  and not carbonatite. This is because Fe, Ca and Mg carbonatite liquids are not stable at low pressure and break down to produce the  $\text{CO}_2$  present in all volcanic emissions. Carbonatite magmas can erupt at Oldoinyo Lengai as a result of their high Na contents; unlike calcic or magnesian carbonate liquids, natro-carbonatite liquids are stable at atmospheric pressure<sup>24</sup>. Therefore, it is the unusual Na content of Oldoinyo Lengai magmas that permits eruption of the carbonatite lavas, not the C content of their source.

Most recent experiments on carbonatite generation use compositions that are enriched in C relative to the current best estimates of average mantle compositions, and show that either high C concentrations (of several thousand p.p.m.) and/or exceedingly small partial-melt fractions are required to produce carbonatitic primary liquids<sup>6</sup>. Furthermore, melting in the presence of C at low partial-melt fraction will produce nephelinitic primary liquids (higher melt fractions tend to dilute the undersaturated melts with melts of more basaltic compositions<sup>8</sup>), consistent with the nephelinitic lavas found at Oldoinyo Lengai. Given that the maximum C concentration of

the Oldoinyo Lengai mantle source is 300 p.p.m., the compromise between C concentration and partial-melt fraction that will still produce an undersaturated primary liquid<sup>8</sup> allows us to constrain the maximum partial-melt fraction at Oldoinyo Lengai to be  $\leq 0.3\%$ . Therefore, rather than highly unusual C-rich fluids in the mantle, or carbonated peridotite, it seems likely that the melting regime in the extensional East African Rift is what results in the low partial-melt fraction liquids of nephelinitic composition. These melts undergo further differentiation and liquid immiscibility in the shallow crust such that  $\sim 1\%$  of the mantle C separates into a natro-carbonatite phase—into which incompatible elements preferentially partition. The remaining C degasses as  $\text{CO}_2$ , as at all other volcanoes. The source for all Oldoinyo Lengai magmas (nephelinites and carbonatites) is ultimately the upper mantle, which is indistinguishable from that under mid-ocean ridges in terms of the isotopic composition of volatiles and lithophile elements.

On a global scale, our results imply that the regions of upper mantle beneath mid-ocean ridges<sup>17</sup>, continental North America<sup>25</sup> and the East African Rift are identical in their volatile abundances and isotopic compositions. Despite small differences in some trace gases (for example Xe isotopes in MORBs versus continental well gases<sup>26</sup>), the upper mantle appears to be a uniform and homogenous geochemical reservoir of  $\text{CO}_2$  and other gases ( $\text{N}_2$ , He and Ar) below both continents and oceans.

## METHODS SUMMARY

All gas samples were collected following methods described in ref. 10. We collected gas samples for gas chemistry and N isotopes in  $\sim 250$ -ml evacuated glass flasks equipped with Teflon stopcocks. Before sample collection, we filled the flasks with  $\sim 50$  ml NaOH solution to absorb the acid gases ( $\text{CO}_2$ ,  $\text{SO}_2$ ,  $\text{H}_2\text{S}$ , HCl and HF) and inert gases ( $\text{N}_2$ ,  $\text{H}_2$ ,  $\text{O}_2$ ,  $\text{CH}_4$ , CO and noble gases) filled the headspace. Samples for determination of He, Ar and C isotopes and  $\text{CO}_2/{}^3\text{He}$  ratios were collected at the same time and from the same locations using evacuated alkaline-resistant-glass flasks with high-vacuum stopcocks. Samples OLD-1 and OLD-2 came from TB49, a degassing hornito (a cone several metres high made of previously erupted lava). OLD-3 and OLD-4 were collected from a nearby hole (2 m in diameter) in an old lava flow. During sampling of OLD-1, OLD-2 and OLD-3, on 4 July 2005, eruption from a vent  $\sim 50$  m away was continuous and lavas were emitted at a rate of  $\sim 0.3 \text{ m}^3 \text{ s}^{-1}$ , flowing at a maximum speed of  $2 \text{ m s}^{-1}$  over the flat crater floor. During sampling of OLD-4, on 5 July 2005, the eruption was less vigorous. The volcano was under increasing gas pressure owing to the rise of magma, and atmospheric contamination, which is always a problem when collecting volcanic gases, was limited. At the University of New Mexico, the headspaces of the gas samples were analysed by gas chromatography (using thermal-conductivity and flame-ionization detectors). From the headspace, gas aliquots were taken for N isotope analyses using an isotope-ratio mass spectrometer equipped with a gas bench<sup>11</sup>. At the Scripps Institution of Oceanography, He and Ar isotopes, C isotopes on  $\text{CO}_2$  and  $\text{CO}_2/{}^3\text{He}$  ratios were measured by a combination of noble-gas and stable-isotope mass spectrometry, and gas manometry following ref. 12.

Received 22 February 2008; accepted 12 March 2009.

1. Dawson, J. B. Sodium carbonatite lavas from Oldoinyo Lengai, Tanganyika. *Nature* **195**, 1075–1076 (1962).
2. Pik, R., Marty, B. & Hilton, D. R. How many plumes in Africa? The geochemical point of view. *Chem. Geol.* **226**, 100–114 (2006).
3. Keller, J. & Spettel, B. in *Carbonatite Volcanism: Oldoinyo Lengai and the Petrogenesis of Natrocarbonatites* (eds Bell, K. & Keller, J.) 70–86 (IAVCEI Proc. Volcanol. Vol. 4, Springer, 1995).
4. Bell, K. & Dawson, J. B. in *Carbonatite Volcanism: Oldoinyo Lengai and the Petrogenesis of Natrocarbonatites* (eds Bell, K. & Keller, J.) 100–112 (IAVCEI Proc. Volcanol. Vol. 4, Springer, 1995).
5. Peterson, T. D. Peralkaline nephelinites II. Low pressure fractionation and the hypersodic lavas of Oldoinyo Lengai. *Contrib. Mineral. Petrol.* **102**, 336–346 (1989).
6. Dasgupta, R. & Hirschmann, M. M. Melting in the Earth's deep upper mantle caused by carbon dioxide. *Nature* **440**, 659–662 (2006).
7. Wyllie, P. J. in *Carbonatites: Genesis and Evolution* (ed. Bell, K.) 500–540 (Unwin Hyman, 1989).
8. Dasgupta, R., Hirschmann, M. M. & Smith, N. D. Partial melting experiments of peridotite +  $\text{CO}_2$  at 3 GPa and genesis of alkalic ocean island basalts. *J. Petrol.* **48**, 2093–2124 (2007).
9. Brantley, S. L. & Koepnick, K. W. Measured carbon dioxide emissions from Oldoinyo Lengai and the skewed distribution of passive volcanic fluxes. *Geology* **23**, 933–936 (1995).



10. Hilton, D. R., Fischer, T. P. & Marty, B. in *Noble Gases in Geochemistry and Cosmochemistry* (eds Porcelli, D., Ballentine, C. & Wieler, R.) 319–362 (Rev. Mineral. Geochem. Vol. 47, Mineralogical Society of America, 2002).
11. Fischer, T. P. *et al.* Subduction and recycling of nitrogen along the Central American margin. *Science* **297**, 1154–1157 (2002).
12. Shaw, A. M., Hilton, D. R., Fischer, T. P., Walker, J. A. & Alvarado, G. Contrasting He–C relationships in Nicaragua and Costa Rica: insights into C cycling through subduction zones. *Earth Planet. Sci. Lett.* **214**, 499–513 (2003).
13. Graham, D. W. in *Noble Gases in Geochemistry and Cosmochemistry* (eds Porcelli, D., Ballentine, C. & Wieler, R.) 247–317 (Rev. Mineral. Geochem. Vol. 47, Mineralogical Society of America, 2002).
14. Marty, B. & Zimmermann, L. Volatiles (He, C, N, Ar) in mid-ocean ridge basalts: assessment of shallow-level fractionation and characterization of source composition. *Geochim. Cosmochim. Acta* **63**, 3619–3633 (1999).
15. Braeuer, K., Kaempfer, H., Niedermann, S., Strauch, G. & Weise, S. M. Evidence for a nitrogen flux directly derived from the European subcontinental mantle in the Western Eger Rift, Central Europe. *Geochim. Cosmochim. Acta* **68**, 4935–4947 (2004).
16. Sano, Y., Takahata, N., Nishio, Y., Fischer, T. P. & Williams, S. N. Volcanic flux of nitrogen from the Earth. *Chem. Geol.* **171**, 263–271 (2001).
17. Burnard, P. G., Graham, D. W. & Turner, G. Vesicle specific noble gas analyses of “popping rock”: implications for primordial noble gases in earth. *Science* **276**, 568–571 (1997).
18. Gilfillan, S. M. V. *et al.* The noble gas geochemistry of natural CO<sub>2</sub> gas reservoirs from the Colorado Plateau and Rocky Mountain provinces, USA. *Geochim. Cosmochim. Acta* **72**, 1174–1198 (2008).
19. Marty, B. & Tolstikhin, I. N. CO<sub>2</sub> fluxes from mid-ocean ridges, arcs and plumes. *Chem. Geol.* **145**, 233–248 (1998).
20. Burnard, P. & Toplis, M. Helium solubility in carbonate liquids; potential for generating high <sup>3</sup>He/U mantle. *Geochim. Cosmochim. Acta* **69** (Suppl. 10), 433 (2005).
21. Jambon, A., Weber, H. & Braun, O. Solubility of He, Ne, Ar, Kr, and Xe in a basalt melt in the range 1250–1600 °C. Geochemical implications. *Geochim. Cosmochim. Acta* **50**, 401–408 (1986).
22. Cartigny, P., Pineau, F., Aubaud, C. & Javoy, M. Towards a consistent mantle carbon flux estimate; insights from volatile systematics (H<sub>2</sub>O/Ce, δD, CO<sub>2</sub>/Nb) in the North Atlantic mantle (14° N and 34° N). *Earth Planet. Sci. Lett.* **265**, 672–685 (2008).
23. Saal, A., Hauri, E. H., Langmuir, C. H. & Perfit, M. Vapor undersaturation in primitive mid-ocean ridge basalt and the volatile content of the Earth's upper mantle. *Nature* **419**, 451–455 (2002).
24. Genge, M. J., Jones, A. P. & Price, G. D. An infrared and Raman study of carbonate glasses: implications for the structure of carbonatite magmas. *Geochim. Cosmochim. Acta* **59**, 927–937 (1995).
25. Holland, G. & Ballentine, C. J. Seawater subduction controls the heavy noble gas composition of the mantle. *Nature* **441**, 186–191 (2006).
26. Pepin, R. O. & Porcelli, D. Xenon isotope systematics, giant impacts, and mantle degassing on the early Earth. *Earth Planet. Sci. Lett.* **250**, 470–485 (2006).
27. Peeters, F. *et al.* Improving noble gas based paleoclimate reconstruction and groundwater dating using <sup>20</sup>Ne/<sup>22</sup>Ne ratios. *Geochim. Cosmochim. Acta* **67**, 587–600 (2002).
28. Sano, Y. & Marty, B. Origin of carbon in fumarolic gas from island arcs. *Chem. Geol.* **119**, 265–274 (1995).
29. Burnard, P. The bubble-by-bubble volatile evolution of two mid-ocean ridge basalts. *Earth Planet. Sci. Lett.* **174**, 199–211 (1999).
30. Porcelli, D. & Ballentine, C. J. in *Noble Gases in Geochemistry and Cosmochemistry* (eds Porcelli, D., Ballentine, C. & Wieler, R.) 412–480 (Rev. Mineral. Geochem. Vol. 47, Mineralogical Society of America, 2002).
31. Newell, D. L., Crossey, L. J., Karlstrom, K. E., Fischer, T. P. & Hilton, D. R. Continental-scale links between the mantle and groundwater systems of the Western US: evidence from travertine springs and regional He isotope data. *GSA Today* **15**, 4–10 (2005).
32. Giggensbach, W. F. in *Monitoring and Mitigation of Volcano Hazards* (eds Scarpa, R. & Tilling, R.) 221–256 (Springer, 1996).
33. Dixon, J. E., Stolper, E. M. & Holloway, J. R. An experimental study of water and carbon dioxide in mid-ocean ridge basaltic liquids. Part I: Calibration and solubility models. *J. Petrol.* **36**, 1607–1631 (1995).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We would like to thank CNRS-INSU for financial support, the French Embassy in Dar Es Salaam and The University of New Mexico Research Allocation Committee for support of field work, the US National Science Foundation for analytical support at the Scripps Institution of Oceanography (EAR-0439122) and the University of New Mexico (EAR-0537618, EAR-0827352). We thank the Tanzania Commission for Science and Technology for granting research permission (2005-217-NA-2005-74) for this study. We thank C. Ballentine for a thorough review that greatly helped to improve this paper.

**Author Contributions** T.P.F. collected gas samples for chemical and isotopic analyses and analysed gas chemistry and N isotopes; P.B. and B.M. designed the study, led the field expedition to Oldoinyo Lengai and obtained funding for the expedition and permits; D.R.H. led the analysis of He, Ar and C isotopes and CO<sub>2</sub>/<sup>3</sup>He ratios; T.P.F., P.B., B.M. and D.R.H. collectively wrote the paper; E.F. analysed noble gases and C isotopes at the Scripps Institution of Oceanography; F.P. helped with sample collection; Z.D.S. supported N isotope analyses at the University of New Mexico; and F.M. helped with logistics in Tanzania, organizing the field expedition and obtaining permits.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to T.P.F. ([fischer@unm.edu](mailto:fischer@unm.edu)).

# The foot of *Homo floresiensis*

W. L. Jungers<sup>1</sup>, W. E. H. Harcourt-Smith<sup>2</sup>, R. E. Wunderlich<sup>3</sup>, M. W. Tocheri<sup>4</sup>, S. G. Larson<sup>1</sup>, T. Sutikna<sup>5</sup>, Rhokus Awe Due<sup>5</sup> & M. J. Morwood<sup>6</sup>

*Homo floresiensis* is an endemic hominin species that occupied Liang Bua, a limestone cave on Flores in eastern Indonesia, during the Late Pleistocene epoch<sup>1,2</sup>. The skeleton of the type specimen (LB1) of *H. floresiensis* includes a relatively complete left foot and parts of the right foot<sup>3</sup>. These feet provide insights into the evolution of bipedalism and, together with the rest of the skeleton, have implications for hominin dispersal events into Asia. Here we show that LB1's foot is exceptionally long relative to the femur and tibia, proportions never before documented in hominins but seen in some African apes. Although the metatarsal robusticity sequence is human-like and the hallux is fully adducted, other intrinsic proportions and pedal features are more ape-like. The postcranial anatomy of *H. floresiensis* is that of a biped<sup>1-3</sup>, but the unique lower-limb proportions and surprising combination of derived and primitive pedal morphologies suggest kinematic and biomechanical differences from modern human gait. Therefore, LB1 offers the most complete glimpse of a bipedal hominin foot that lacks the full suite of derived features characteristic of modern humans and whose mosaic design may be primitive for the genus *Homo*. These new findings raise the possibility that the ancestor of *H. floresiensis* was not *Homo erectus* but instead some other, more primitive, hominin whose dispersal into southeast Asia is still undocumented.

Bipedalism is the derived hallmark of hominins<sup>4,5</sup>, yet our understanding of its origins and subsequent evolution is incomplete. Associated postcranial elements are rare in the early part of the hominin fossil record<sup>5,6</sup>, and relatively complete feet are rarer still. Although LB1 has been dated to the end of the Pleistocene<sup>1,3</sup>, its feet shed new light on the evolution of bipedal gait and have important implications for hominin biogeography. The reassembled left foot is shown in Fig. 1 in both medial and dorsal view next to the associated right tibia and left femur.

It is possible to estimate the full length of LB1's foot (to the tip of the second digit) from the assembled partial foot, assuming a modern human template and using a regression analysis based on a large sample of complete modern human feet (Supplementary Fig. 1). The estimated skeletal length of LB1's left foot is 191 mm (95% prediction interval, 184–199 mm), indicating a fleshy-foot length<sup>7</sup> of 196 mm (189–204 mm). As Fig. 1 shows, the foot of LB1 is very long relative to the femur and tibia. The ratio of fleshy-foot length to maximum femur length (280 mm in LB1; ref. 2) is 0.7 (0.68–0.73). The relative foot length in LB1 far exceeds the upper limits for modern humans of either average or short stature (Table 1) and instead overlaps with bonobos (*Pan paniscus*). It is also relatively longer than the estimate for *Australopithecus afarensis* (AL 288-1), which was based in part on an Early Pleistocene foot from Olduvai Gorge, Tanzania (OH 8)<sup>7</sup>. Because it is the distal-most segment in the lower-limb link system, a hominin foot of this relative length requires

compensatory kinematic adjustments at ankle, knee and/or hip joints to permit adequate clearance of the toes in the swing phase of walking and running (unless the toes were curled)<sup>8</sup>, and it probably places limits on running speed<sup>9</sup>. The relatively high foot-to-femur ratio, not unlike the high humerofemoral index<sup>2</sup>, is driven primarily by an exceptionally short hindlimb (Supplementary Table 1).

The metatarsal robusticity sequence observed in LB1 follows the most common human formula, I > V > IV > III > II (refs 3, 10). The orientation of the virtually flat and mutually articulating surfaces of the entocuneiform and hallucal metatarsal indicate that the big toe was fully adducted and in line with the other metatarsals. However, in comparison with the lesser metatarsals, the hallucal metatarsal is short; its relative length falls below the range observed in modern humans but is within the range characteristic of chimpanzees (Fig. 2a). Dorsal extensions of the articular surfaces are evident on all metatarsal heads, and the distal hallucal metatarsal is squared off as in modern humans but not in apes, *A. afarensis*, *Paranthropus robustus*<sup>11</sup> or early *Homo* from Dmanisi, Georgia<sup>6</sup>. Sesamoid grooves are not evident on the distal joint surface of the hallucal metatarsal of LB1. Modern human-like dorsiflexion and stability at the metatarsophalangeal joints late in stance phase are implied, but load-sharing among the digits at toe-off was probably impacted by the relatively



**Figure 1** | The foot and long bones of the type specimen of *H. floresiensis*. Medial (top) and dorsal views of the reassembled left foot of LB1, shown here with its associated right tibia and left femur.

<sup>1</sup>Department of Anatomical Sciences, Stony Brook University Medical Center, Stony Brook, New York 11794-8081, USA. <sup>2</sup>Division of Paleontology, American Museum of Natural History, New York, New York 10024, USA. <sup>3</sup>Department of Biological Sciences, James Madison University, Harrisonburg, Virginia 22807, USA. <sup>4</sup>Humans Origins Program, Department of Anthropology, National Museum of Natural History, Smithsonian Institution, Washington DC 20013-7012, USA. <sup>5</sup>National Research and Development Centre for Archaeology, Jakarta 12001, Indonesia. <sup>6</sup>School of Earth and Environmental Sciences, University of Wollongong, Wollongong, New South Wales 2522, Australia.



**Table 1 | Relative fleshy-foot length in LB1, humans and apes\***

| Species                        | N  | Mean | Standard deviation | Range      |
|--------------------------------|----|------|--------------------|------------|
| <i>Homo sapiens</i> †          | 34 | 54.2 | 2.5                | 49.3–58.9  |
| <i>Homo sapiens</i> (pygmy)‡   | 10 | 54.5 | 2.1                | 50.7–56.7  |
| <i>Homo floresiensis</i> (LB1) | 1  | 70.0 | —                  | 67.5–72.9§ |
| <i>Pan paniscus</i>            | 8  | 73.9 | 3.2                | 68.4–77.2  |
| <i>Pan troglodytes</i>         | 36 | 82.5 | 4.9                | 72.3–95.9  |

N, sample size.

\*  $100 \times (\text{fleshy-foot length}/\text{femur length})$ .

† Data from ref. 7 (Cleveland Museum of Natural History).

‡ Sample includes small-bodied individuals from Africa, the Andaman Islands and South America in the collections of the Natural History Museum, London, and the American Museum of Natural History, New York.

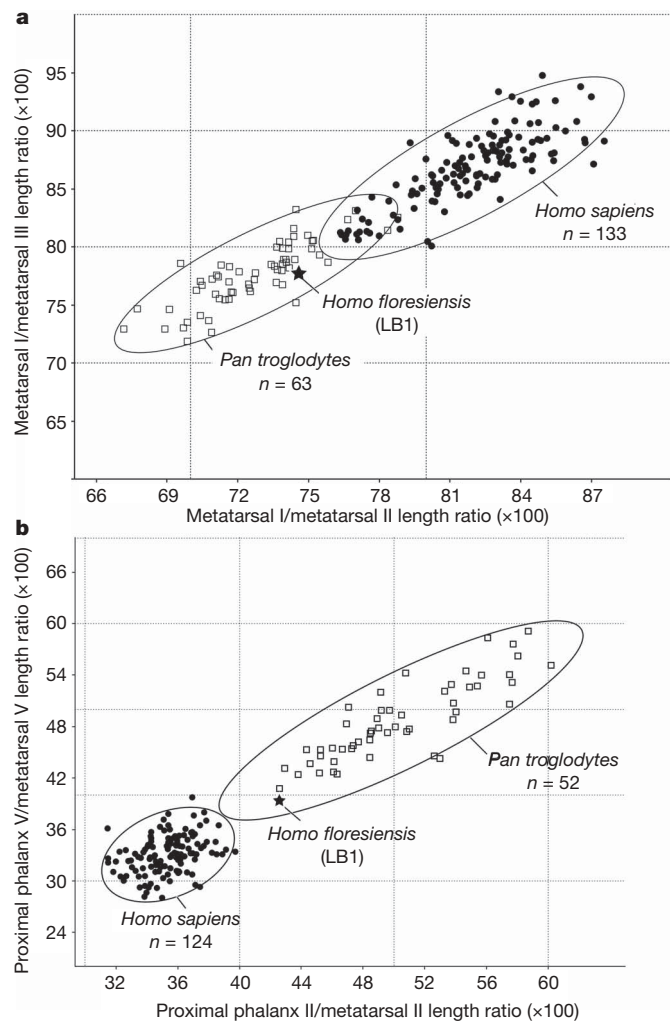
§ 95% prediction interval.

short hallux and long foot of LB1 (ref. 12). At  $5\text{--}7^\circ$  (lateral), the degree of torsion of the distal tibia of LB1 relative to its proximal platform is at the lower end of the observed human range; despite the short hallux, therefore, there is no indication that the foot of LB1 was toed in<sup>13</sup>.

The forefoot of LB1 is disproportionately long in comparison with modern humans. Relative to tarsal length (proximal talus to distal entocuneiform), the lesser metatarsals are long and fall outside the observed limits of our modern human sample (Supplementary Fig. 2). The proximal pedal phalanges also contribute to the long forefoot of LB1. As a percentage of corresponding metatarsal lengths, the proximal pedal phalanges of LB1 are longer than those observed in modern humans, and instead more closely resemble those in some chimpanzees (Fig. 2b). The proximal pedal phalanges of LB1 have robust bases and midshafts, and they lack the hour-glass shape characteristic of modern humans. They are also moderately curved, with included angles ranging from  $16.8^\circ$  to  $26.8^\circ$ ; the second value is beyond the range documented for modern humans and resembles some australopithecines<sup>14</sup>. Modern humans have short, straight toes<sup>15</sup>; short toes reduce the torques and bending stresses of ground-reaction forces in bipedalism and serve to reduce mechanical work in running<sup>16</sup>. The proximal pedal phalanges of LB1 lack this derived design feature, and thus probably experienced habitually high stresses (as further suggested by the presence of osteophytes on several of the phalanges)<sup>3</sup>. Their exceptionally high levels of robusticity also indicate that LB1 was unshod (Supplementary Table 2).

Of all the tarsal bones of LB1, the navicular is the most primitive in shape (Fig. 3a). The navicular tuberosity is large and projecting, and the lateral edge of the bone is pinched, producing a wedge-like effect seen in australopithecines and great apes but not in modern humans or OH 8 (ref. 17). The entocuneiform facet is well separated from the enlarged tuberosity, and there is no contact facet for the cuboid (a variable feature in modern humans). As in great apes and australopithecines, the medial longitudinal arch was probably weakly developed or absent in LB1 (refs 15, 17) and the navicular tuberosity was probably weight-bearing<sup>18</sup>. LB1's navicular affinities are corroborated by a statistical analysis of shape by means of geometric morphometrics (Fig. 3b, Supplementary Table 3 and Supplementary Fig. 3). The first two principal components of three-dimensional (Procrustes) shape coordinates reveal complete separation of modern humans from great apes. The left and right naviculars of LB1 and two of the three australopithecine naviculars group with African apes; OH 8 is more similar in shape to modern humans.

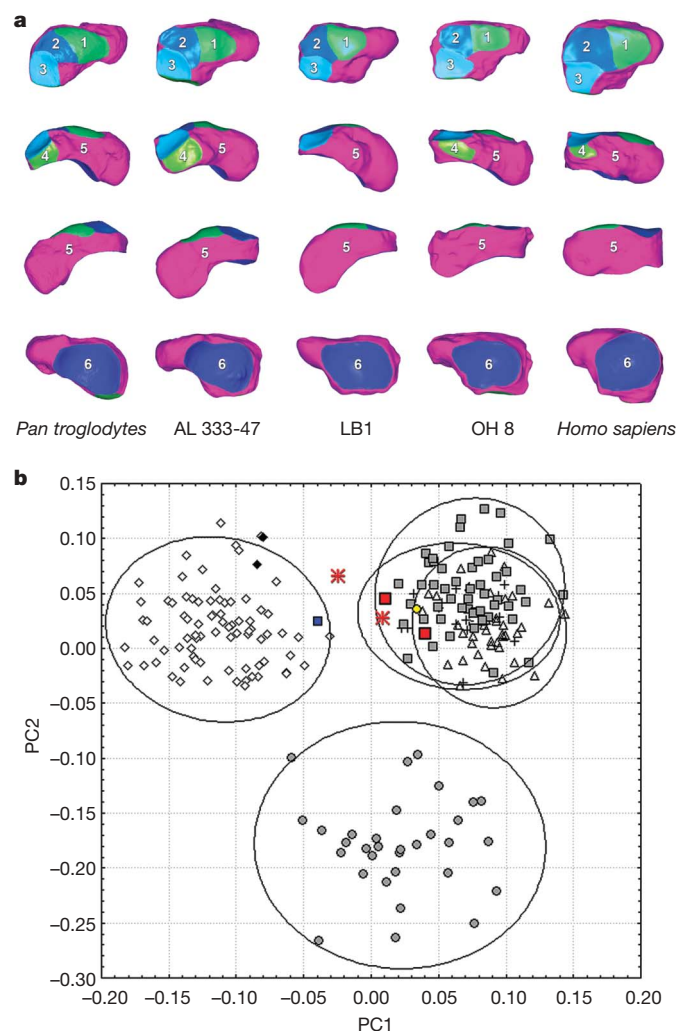
The talus of LB1 (Supplementary Information Fig. 4) is similar to modern humans in some features (for example, the neck angle is  $23^\circ$ , the body is slightly wedge-shaped and the dorsal trochlear groove is shallow with subequal margins)<sup>19</sup>, but it departs towards ape-like morphology with a very low degree of talar head torsion (at  $26^\circ$ , it is two to four standard deviations below mean values for modern humans)<sup>20</sup>. A three-dimensional geometric morphometric analysis shows that LB1's talus falls outside of and between the envelopes for modern humans and African apes; it is 'intermediate' in shape, as are the tali of OH 8 and KNM-ER 1476a (Supplementary Figs 5 and 6).



**Figure 2 | Intrinsic foot proportions in *H. floresiensis*.** **a**, Length of the hallux metatarsal relative to lengths of metatarsals II and III in humans, chimpanzees and LB1 (95% correlation ellipses are indicated for humans and chimpanzees). The human sample includes 124 individuals from the Indian subcontinent, five African pygmies and sex-specific means for Europeans and Japanese. *n*, sample size. **b**, Proximal phalangeal length relative to metatarsal length for digits II and V. The more complete right metatarsal V is substituted here for the fragmentary left one. Following a human template, the longest proximal phalanx (LB1/38) was assigned to metatarsal II, and the shortest (LB1/34) to metatarsal V (ref. 3).

The cuboids of LB1 appear less trapezoidal in shape than many modern humans, but both bear an incipient calcaneal process or asymmetrical 'beak' (Fig. 4), which is usually regarded as evidence of a stable lateral longitudinal column<sup>21</sup>. However, expression of this feature is variable in both modern humans<sup>21</sup> and *H. floresiensis*; this surface is nearly flat in a newly discovered right cuboid (LB16) of another individual. In contrast, OH 8 exhibits a very pronounced calcaneal process<sup>17,22</sup>.

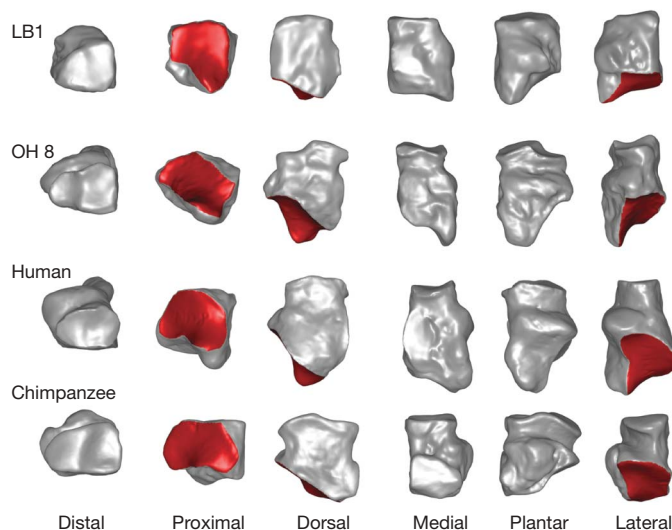
The bony pelvis, lower-limb bones and feet of *H. floresiensis* combine to reveal an unequivocal adaptation to bipedalism<sup>1–3</sup>, yet they also depart in functionally significant ways from the skeletal design of modern humans, with important biomechanical and evolutionary implications. For instance, because the foot lacks a well-defined medial longitudinal arch, recovery of stored elastic energy would have been limited in comparison with modern humans, especially during running, when mass-spring mechanics replace the pendular mechanics of walking<sup>15</sup>. With a short hallux, relatively long lateral rays and a weight-bearing navicular tuberosity, pressure transfer through the foot in support phase and at toe-off also probably differed from modern humans. Kinematic differences related to clearance of the foot



**Figure 3 | The comparative anatomy and morphometrics of the navicular bone in *H. floresiensis*.** **a**, Laser scans of naviculars from *P. troglodytes*, *A. afarensis* (AL 333-47), *H. floresiensis* (LB1), OH 8 and *H. sapiens* (scaled to approximately similar size and shown as from the right side). From top to bottom, the rows show distal, plantar, dorsal and proximal views. Articular facets are indicated: entocuneiform (green, 1), mesocuneiform (blue, 2), ectocuneiform (light blue, 3), cuboid (light green, 4) and talus (purple, 6); non-articular surface (magenta, 5). **b**, First two principal components of a generalized Procrustes analysis of three-dimensional coordinates (see Supplementary Fig. 4), with 95% confidence ellipses. PC1 accounts for 24.6% of the variance, and PC2 for 19% (*H. sapiens*, open diamonds; *P. troglodytes*, open triangles; *P. paniscus*, crosses; *Gorilla*, grey squares; *Pongo*, grey circles; human pituitary dwarf, black diamonds; OH 8, blue square; *A. afarensis*, red asterisks; South African *Australopithecus*, yellow circle; *H. floresiensis* (left and right sides), red squares). Sample sizes, composition and statistical software are documented in Supplementary Table 3.

in swing phase would have been evident during both walking and running. The foot of *H. floresiensis* was not well-designed for either high-speed or efficient endurance running<sup>15,16</sup>.

Concerning phylogeny, the foot of *H. floresiensis* exhibits a broad array of primitive features that are not seen in modern humans of any body size. Primitive traits are also seen in the LB1 cranium, mandible, brain, shoulder, wrist, pelvis, limb bones and body proportions<sup>1–3,23–26</sup>. It is conceivable that a few of these plesiomorphic traits could have evolved through reversals during ~800 kyr of insular isolation on Flores<sup>1,27,28</sup>, but it is improbable that all of them from head to toe were simply a consequence of 'island dwarfing'. Some modern humans (pygmies) have reduced greatly in body size repeatedly and independently throughout the world, without any evidence of evolutionary reversals to such primitive morphologies and body



**Figure 4 | The comparative anatomy of the cuboid bone in *H. floresiensis*.** High-resolution laser scans of the left cuboid in LB1, OH 8, a human and a chimpanzee (scaled to approximately similar size). The calcaneal facet is highlighted in red. The calcaneal beak is present but weakly expressed in LB1; it is especially pronounced in OH 8.

proportions<sup>29</sup>. Re-evolving short hindlimbs, long lateral toes, a short hallux and a flat foot would impair locomotor performance, and no known human skeletal pathology fully recapitulates this ancient body design. The comparative and functional anatomical evidence of the foot (and much of the rest of the skeleton) suggests that *H. floresiensis* possesses many characteristics that may be primitive for the genus *Homo*. It follows that if these features are primitive retentions, then *H. floresiensis* could be a descendant of a primitive hominin that established a presence in Asia either alongside or at a different time than *H. erectus sensu stricto*<sup>30</sup>.

Received 14 January; accepted 17 March 2009.

1. Brown, P. et al. A new small-bodied hominin from the Late Pleistocene of Flores, Indonesia. *Nature* **431**, 1055–1061 (2004).
2. Morwood, M. J. et al. Further evidence for small-bodied hominins from the Late Pleistocene of Flores, Indonesia. *Nature* **437**, 1012–1017 (2005).
3. Jungers, W. L. et al. Descriptions of the lower limb skeleton of *Homo floresiensis*. *J. Hum. Evol.* doi:10.1016/j.jhevol.2008.08.014 (in the press).
4. Zollikofer, C. P. E. et al. Virtual cranial reconstruction of *Sahelanthropus tchadensis*. *Nature* **434**, 755–759 (2005).
5. Richmond, B. G. & Jungers, W. L. *Orrorin tugenensis* femoral morphology and the evolution of hominin bipedalism. *Science* **319**, 1662–1665 (2008).
6. Lordkipanidze, D. et al. Postcranial evidence from early *Homo* from Dmanisi, Georgia. *Nature* **449**, 305–310 (2007).
7. White, T. D. & Suwa, G. Hominid footprints at Laetoli: facts and interpretations. *Am. J. Phys. Anthropol.* **72**, 485–514 (1987).
8. Winter, D. A. Foot trajectory in human gait: a precise and multifactorial motor control task. *Phys. Ther.* **72**, 45–53 (1992).
9. Jungers, W. L. & Stern, J. T. Jr. Body proportions, skeletal allometry and locomotion in the Hadar hominids: a reply to Wolpoff. *J. Hum. Evol.* **12**, 673–684 (1983).
10. Archibald, J. D., Lovejoy, C. O. & Heiple, K. G. Implications of relative robusticity in the Olduvai metatarsus. *Am. J. Phys. Anthropol.* **37**, 93–96 (1972).
11. Susman, R. L. & Brain, T. M. New first metatarsal (SKX 5017) from Swartkrans and the gait of *Paranthropus robustus*. *Am. J. Phys. Anthropol.* **77**, 7–15 (1988).
12. Rodgers, M. M. & Cavanagh, P. R. Pressure distribution in Morton's foot structure. *Med. Sci. Sports Exerc.* **21**, 23–28 (1989).
13. Wallace, I. J. et al. The bipedalism of the Dmanisi hominins: pigeon-toed early *Homo*? *Am. J. Phys. Anthropol.* **136**, 375–378 (2008).
14. Susman, R. L., Stern, J. T. & Jungers, W. L. Arboreality and bipedality in the Hadar hominids. *Folia Primatol. (Basel)* **43**, 113–156 (1984).
15. Bramble, D. M. & Lieberman, D. E. Endurance running and the evolution of *Homo*. *Nature* **432**, 345–352 (2004).
16. Rolian, C. et al. Walking, running and the evolution of short toes in humans. *J. Exp. Biol.* **212**, 713–721 (2009).
17. Harcourt-Smith, W. E. H. & Aiello, L. C. Fossils, feet, and the evolution of bipedal locomotion. *J. Anat.* **204**, 403–416 (2004).
18. Vereecke, E. et al. Dynamic plantar pressure distribution during terrestrial locomotion of bonobos (*Pan paniscus*). *Am. J. Phys. Anthropol.* **120**, 373–383 (2003).

19. Gebo, D. L. & Schwartz, G. T. Foot bones from Orno: implications for hominid evolution. *Am. J. Phys. Anthropol.* **129**, 499–511 (2006).
20. Rhoads, J. G. & Trinkaus, E. Morphometrics of the Neandertal talus. *Am. J. Phys. Anthropol.* **46**, 29–44 (1977).
21. Bojsen-Møller, F. Calcaneocuboid joint and stability of the longitudinal arch of the foot at high and low gear push off. *J. Anat.* **129**, 165–176 (1979).
22. Lewis, O. J. *Functional Morphology of the Evolving Hand and Foot* 328–333 (Oxford Univ. Press, 1989).
23. Argue, D., Donlon, D., Groves, C. & Wright, R. *Homo floresiensis*: microcephalic, pygmoid, *Australopithecus* or *Homo*? *J. Hum. Evol.* **51**, 360–374 (2006).
24. Falk, D. *et al.* Brain shape in human microcephalics and *Homo floresiensis*. *Proc. Natl Acad. Sci. USA* **104**, 2513–2518 (2007).
25. Larson, S. G. *et al.* *Homo floresiensis* and the evolution of the hominin shoulder. *J. Hum. Evol.* **53**, 718–731 (2007).
26. Tocheri, M. W. *et al.* The primitive wrist of *Homo floresiensis* and its implications for hominin evolution. *Science* **317**, 1743–1745 (2007).
27. Morwood, M. J., O'Sullivan, P. B., Aziz, F. & Raza, A. Fission-track age of stone tools and fossils on the east Indonesian island of Flores. *Nature* **392**, 173–176 (1998).
28. Lyras, G. S. *et al.* The origin of *Homo floresiensis* and its relation to evolutionary processes under isolation. *Anthropol. Sci.* advance online publication doi:10.1537/ase.080411 (1 August 2008).
29. Migliano, A. B., Vinicius, L. & Lahr, M. M. Life history trade-offs explain the evolution of human pygmies. *Proc. Natl Acad. Sci. USA* **104**, 20216–20219 (2007).
30. Dennell, R. & Roebroeks, W. An Asian perspective on early human dispersal from Africa. *Nature* **438**, 1099–1104 (2005).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This work was supported by grants from the Australian Research Council, the National Geographic Society, the Wenner-Gren Foundation for Anthropological Research, the Wellcome Trust and the Leakey Foundation. We also thank our colleagues Jatmiko, E. Wahyu Saptomo and T. Djubiantono for their encouragement and assistance in the excavations.

**Author Contributions** W.L.J., W.E.H.H.-S., R.E.W., M.W.T., M.J.M. and S.G.L. undertook the analyses of the pedal remains and wrote the paper; T.S. and R.A.D. were responsible for the Liang Bua excavations and assisted with data collection and analyses.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to W.L.J. ([william.jungers@stonybrook.edu](mailto:william.jungers@stonybrook.edu)).



# Insular dwarfism in hippos and a model for brain size reduction in *Homo floresiensis*

Eleanor M. Weston<sup>1</sup> & Adrian M. Lister<sup>1</sup>

Body size reduction in mammals is usually associated with only moderate brain size reduction, because the brain and sensory organs complete their growth before the rest of the body during ontogeny<sup>1,2</sup>. On this basis, 'phyletic dwarfs' are predicted to have a greater relative brain size than 'phyletic giants'<sup>1,3</sup>. However, this trend has been questioned in the special case of dwarfism of mammals on islands<sup>4</sup>. Here we show that the endocranial capacities of extinct dwarf species of hippopotamus from Madagascar are up to 30% smaller than those of a mainland African ancestor scaled to equivalent body mass. These results show that brain size reduction is much greater than predicted from an intraspecific 'late ontogenetic' model of dwarfism in which brain size scales to body size with an exponent of 0.35. The nature of the proportional change or grade shift<sup>2,5</sup> observed here indicates that selective pressures on brain size are potentially independent of those on body size. This study demonstrates empirically that it is mechanistically possible for dwarf mammals on islands to evolve significantly smaller brains than would be predicted from a model of dwarfing based on the intraspecific scaling of the mainland ancestor. Our findings challenge current understanding of brain–body allometric relationships in mammals and suggest that the process of dwarfism could in principle explain small brain size, a factor relevant to the interpretation of the small-brained hominin found on the Island of Flores, Indonesia<sup>6</sup>.

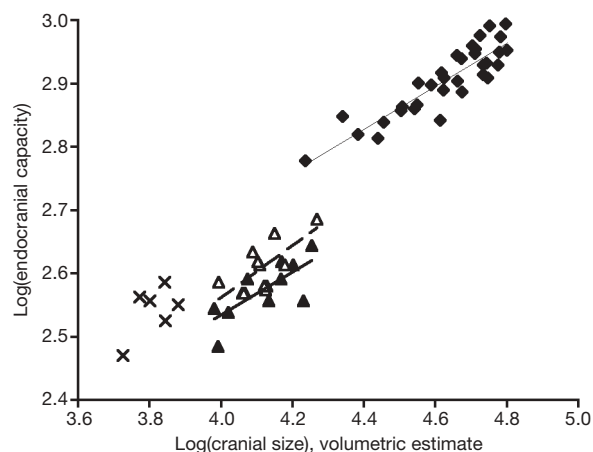
Brain tissue is energetically expensive, and it has been suggested that a decrease in brain volume may be advantageous to an animal's survival under the environmental conditions associated with islands<sup>4,7,8</sup>. This phenomenon, if real, can be difficult to test because not only are most examples of dwarf island mammals extinct and known only from incomplete subfossil material, but knowledge of the founding ancestor can also be difficult to ascertain. The strongest case previously documented is that of a fossil bovid *Myotragus*, isolated on the Mediterranean island of Majorca for more than 5 Myr (refs 4, 9). The brain mass of *Myotragus* was decreased by up to 50% relative to those of extant bovids of equivalent body mass<sup>4</sup>. However, a link between the relatively small brain of *Myotragus* and the process of dwarfism has been disputed<sup>10,11</sup> in the light of the length of the bovid's isolation and its uncertain ancestry<sup>9</sup>.

One aspect of the continuing debate<sup>12–14</sup> over whether the small-brained hominin discovered on the Island of Flores, Indonesia<sup>6</sup>, evolved by insular dwarfism centres on scaling exponents between brain and body size<sup>10,11</sup>. For predicting the brain size of a mammal at a smaller body size, these 'intraspecific' scaling models are the most appropriate to accommodate the likely correlated effects of body size adjustment on brain size in closely related species<sup>10,15,16</sup> (see Supplementary Discussion). However, here we show that two species of extinct dwarf hippopotamus do not correspond to such dwarfing models because even though the same scaling exponents relating brain size to body size apply, large intercept shifts (grade shifts<sup>2,5</sup>)

distinguish these 'phyletic dwarfs' from their mainland ancestor (Fig. 1, Supplementary Table 1 and Supplementary Discussion).

To model the dwarfing process, we took the extant large *H. amphibius*, a generalized representative of the genus *Hippopotamus*, to be the probable ancestor of two recently extinct dwarfed hippos, *H. lemerlei* and *H. madagascariensis*, from the island of Madagascar<sup>17</sup> (see Supplementary Discussion). These species can be unequivocally separated taxonomically<sup>17</sup> but both share apomorphies with the modern genus *Hippopotamus*<sup>18</sup>. The modern pygmy hippopotamus *Choeropsis liberiensis* represents a lineage distinct from all other hippopotamids, diverging from its closest relative before 5 Myr ago (ref. 18). Madagascar is a large island of diverse habitats that has prehistorically supported up to three species of hippopotamus<sup>19</sup>, whose times of dispersal to Madagascar are unknown but whose remains persist to within the past 6,000 years<sup>20</sup>.

Brain–body scaling trends in hippos can be inferred from the relationship of brain to cranial size, because the relation between a volumetric estimate of cranial size and a postcranial estimate of body mass across the two living hippos and the Madagascan dwarf species is approximately isometric (Supplementary Figs 3–7, Supplementary Table 4 and Supplementary Discussion). Predictions of decrease in brain size given in Table 1 are thus based on observed cranial size data and not on estimates of body mass (the latter are provided for contextual purposes only). In Table 1 we evaluate predictions of brain–body scaling in insular dwarf hippos and a dwarf elephant by considering



**Figure 1 | Relationship between brain size and cranial size for an intraspecific 'late ontogenetic' model of dwarfing.** Major axis slopes (95% confidence intervals): thin line, *H. amphibius*, 0.3482 (0.28–0.41); thick line, *H. lemerlei*, 0.369 (0.15–0.63); dashed line, *H. madagascariensis*, 0.4587 (0.20–0.79). Filled diamonds, *H. amphibius* ( $n = 33$ ); filled triangles, *H. lemerlei* ( $n = 12$ ); open triangles, *H. madagascariensis* ( $n = 12$ ); crosses, *C. liberiensis* ( $n = 6$ ). For statistical comparisons see Supplementary Table 1.

<sup>1</sup>Department of Palaeontology, Natural History Museum, London SW7 5BD, UK.

**Table 1 | Estimates of brain size based on dwarfing models**

| Species   | Body mass* (kg)     | Endocranial capacity (cm <sup>3</sup> ) | Prediction of endocranial capacity† |                      |
|---|---------------------|---|-------------------------------------|----------------------|
|   |                     |   | Late ontogenetic scaling‡           | Ontogenetic scaling§ |
| <i>Hippopotamus lemerlei</i>                            | 374 (25%)           | 380 (43%) [70%, 83%]                    | 544 (62%)                           | 456 (52%)            |
| <i>Hippopotamus madagascariensis</i>                    | 393 (26%)           | 421 (48%) [76%, 91%]                    | 553 (63%)                           | 465 (53%)            |
| <i>Choeropsis liberiensis</i>                           | 228 (ref. 19) (15%) | 350 (40%) [83%, 105%]                   | 421 (48%)                           | 334 (38%)            |
| <i>Palaeoloxodon falconeri</i> (scaled to cranial size) | 200 (ref. 22) (2%)  | 1,800 (20%) [80%, 125%]                 | 2,250 (25%)                         | 1,440 (16%)          |
| <i>Palaeoloxodon falconeri</i> (scaled to body mass)    | 100 (ref. 23) (1%)  | 1,800 (20%) [100%, 167%]                | 1,800 (20%)                         | 1,080 (12%)          |

Values in parentheses are expressed as a percentage of the original mean value of the mainland ancestor (*H. amphibius* for the Malagasy hippos, and *P. antiquus* for *P. falconeri*) or, for *C. liberiensis*, the larger sister taxon (*H. amphibius*) (see Table 2). For endocranial capacity, numbers in square brackets indicate the observed values as a percentage of those predicted on the basis of the 'late ontogenetic' and 'ontogenetic' scaling models, respectively. For example, the endocranial capacity of *H. lemerlei* is decreased by 30% of the value predicted by the late ontogenetic model (endocranial capacity = 70% of predicted value), whereas the endocranial capacity of *H. madagascariensis* is decreased by 24% of the value predicted by the late ontogenetic model (endocranial capacity = 76% of predicted value). The scaling exponents ( $k$ ) are modelled from *H. amphibius* postnatal cranial data (Supplementary Table 1). The Malagasy hippo body masses are estimated assuming isometry to cranial volume of the ancestor using an adult body mass of 1,495  $\pm$  29.5 kg for *H. amphibius* (see Supplementary Discussion, and Table 2 for limb-bone estimates of body mass). *P. falconeri* brain size reduction is estimated from both cranial size (50-fold smaller than the ancestor<sup>22</sup>) and an estimate of body mass (100-fold smaller<sup>23</sup>) (*P. antiquus* has an approximate average mass of 10 tonnes, estimated from limb-bone data in ref. 25 and P. Davies, unpublished observations).

\* Malagasy hippo estimates based on cranial size.

† Based on intraspecific dwarfing models using cranial size.

‡  $k = 0.35$ : late phase of brain growth in the age range 2–40 years.

§  $k = 0.47$ : early and late phases of brain growth in the age range 0–40 years.

two intraspecific dwarfing models: 'late ontogenetic' and 'ontogenetic'. In mammals that commonly dwarfed on islands, such as hippos, elephants and deer (see Supplementary Discussion), all of which have precocial young, there is generally thought to be a transition from rapid to slow growth of the brain that closely coincides with birth<sup>2,15,16</sup>. In *H. amphibius*, however, the rapid early phase of brain growth continues after birth for up to 2 years (Supplementary Fig. 1). The 'late ontogenetic' exponent (0.35) in this context is calculated from the late, slower phase of postnatal brain growth (2–40 years) and not just from static adult data. This distinction is important, because the 'growth' as opposed to the 'static adult' exponent will best characterize the size adjustments (developmental allometry) associated with the late phase of brain maturation<sup>1</sup>. However, because all late postnatal brain development in mammals typically involves a low exponent value (0.2–0.4 for intraspecific brain–body scaling)<sup>5</sup>, 'late ontogenetic' exponents should be similar to, or only moderately higher than, those derived from static adult data (see Supplementary Discussion). The higher 'ontogenetic' exponent (0.47) is derived from a complete postnatal series from birth to adulthood in which rapid and slow growth phases are combined.

If the endocranial capacity of *H. lemerlei* scaled as (cranial size)<sup>0.35</sup> in accord with the 'late ontogenetic' model (Fig. 1 and Supplementary Table 1), an endocranial capacity of 544 cm<sup>3</sup> would be predicted; that is, 62% of the value of an ancestor four times its cranial size (Tables 1 and 2). The observed mean endocranial capacity of *H. lemerlei* is 380  $\pm$  7.25 cm<sup>3</sup> (mean  $\pm$  s.e.m.; Table 2), 30% smaller than that predicted from the 'late ontogenetic' model (Table 1). The observed mean endocranial capacity of *H. madagascariensis* is 421  $\pm$  11.9 cm<sup>3</sup> (Table 2), 24% smaller than that predicted from the 'late ontogenetic' model (Table 1). The 'ontogenetic' model, in which the endocranial capacity of *H. lemerlei* would scale as (cranial size)<sup>0.47</sup> (Fig. 2 and Supplementary Table 1), predicts a endocranial capacity of 456 cm<sup>3</sup>, 52% of the value of the ancestor, but the observed endocranial capacity is still 17% smaller than that predicted under this model, and in *H. madagascariensis* it is 9% smaller (Table 1).

The Malagasy dwarf hippos do not correspond to either 'intraspecific' scaling model (Figs 1 and 2 and Table 1), indicating that 'insular

dwarfing' does not necessarily comply with predictions based on the ontogenetic scaling of the mainland ancestor. The Malagasy dwarf hippos are roughly equivalent in body mass to a 1–2-year-old *H. amphibius* (353–544 kg), a stage in development that marks the completion of the first rapid phase of brain growth. In contrast, the extant pygmy hippopotamus (*C. liberiensis*), with a body mass one-sixth to one-eighth that of *H. amphibius* (ref. 19 and Table 2), has a mean endocranial capacity of 350  $\pm$  12.7 cm<sup>3</sup>, approximately equal to the values predicted from the 'ontogenetic' scaling model (Table 1) but reduced by 17% relative to the values predicted from the 'late ontogenetic' model (Table 1). The extant pygmy hippopotamus is not an insular dwarf<sup>21</sup> and is phylogenetically more basal than *H. amphibius*, its only extant sister taxon<sup>18</sup> (see Supplementary Discussion). Adult *C. liberiensis* has a body mass equivalent to a 6-month-old *H. amphibius* (see Supplementary Discussion), a stage of development in which rapid early brain growth persists in the hippopotamus. In primates, too, a marked increase in relative brain size has been attributed to the prolongation of the rapid 'prenatal' phase of brain development<sup>3</sup>. In 'phyletic dwarfs' the opposite developmental adjustment, involving a decrease in the duration of rapid early brain growth, could potentially explain a 'grade shift' to a lower brain–body ratio, as demonstrated by the change in ratio of brain to cranial size in Fig. 1 (see Supplementary Discussion).

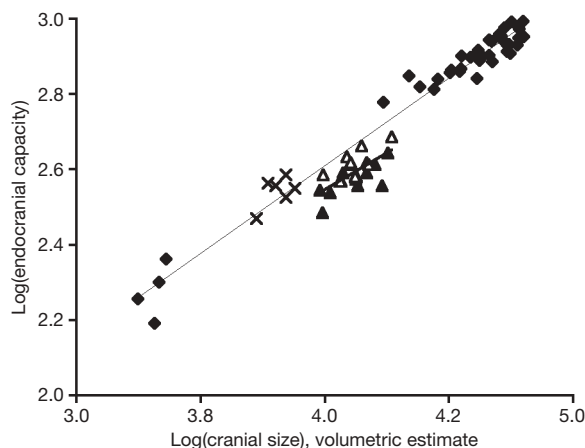
A further example is provided by *Palaeoloxodon falconeri*<sup>22,23</sup>, the smallest insular dwarf elephant from Sicily and Malta, derived from the mainland *P. antiquus*<sup>22,23</sup>. The brain of a modern elephant grows relatively little after birth (the brain of a newborn elephant is 50% of its adult brain weight, whereas that of a common hippopotamus is 22% and that of a human is 25%)<sup>24</sup>, indicating that rapid early brain growth is completed by the end of gestation. The adult *P. falconeri* is estimated to be approximately the size of a newborn modern elephant<sup>23</sup>, and hence brain mass is expected to scale to body mass with a 'late ontogenetic' exponent typical of a precocial mammal. With brain size scaled as (body mass)<sup>0.35</sup>, brain size matches the prediction of the 'late ontogenetic' model with a 100-fold estimated decrease in body mass (Table 1, second entry for *P. falconeri*; see also refs 10, 11). However, estimates of decrease in body mass in these elephants are imprecise,

**Table 2 | Endocranial capacities, cranial sizes and body masses for species in Table 1**

| Species                    | Endocranial capacity (cm <sup>3</sup> ) | Adult cranial size (cm <sup>3</sup> ) | Body mass range (kg)   |
|----------------------------|---|---------------------------------------|------------------------|
| <i>H. amphibius</i>        | 882 $\pm$ 16 (n = 18)                   | 52,533 $\pm$ 1,657 (n = 20)           | 1,210–2,001*           |
| <i>H. lemerlei</i>         | 380 $\pm$ 7.25 (n = 24)                 | 13,298 $\pm$ 810 (n = 12)             | 274–393                |
| <i>H. madagascariensis</i> | 421 $\pm$ 11.9 (n = 12)                 | 13,948 $\pm$ 751 (n = 12)             | 310–642                |
| <i>C. liberiensis</i>      | 350 $\pm$ 12.7 (n = 6)                  | 6,524 $\pm$ 338 (n = 6)               | 180–275 (ref. 19)      |
| <i>P. antiquus</i>         | 9,000 (ref. 22) (n = 1)                 | 311,000 (ref. 22) (n = 1)             | 5,055–17,675 (ref. 25) |
| <i>P. falconeri</i>        | 1,800 (ref. 22) (n = 1)                 | 6,120 (ref. 22) (n = 1)               | 80–200 (refs 22, 23)   |

The endocranial capacity and cranial size (volumetric estimate) data for all species of hippopotamus were collected as part of this study (see Methods). The elephant cranial data are from ref. 22. Observed ranges of adult body mass are given for the extant species (see Supplementary Discussion), and estimates based on postcrania are given for the extinct species (see Supplementary Table 4 for Malagasy hippo data, and Supplementary Discussion). Results are given as means  $\pm$  s.e.m.

\* See Supplementary Discussion.



**Figure 2 | Relationship between brain size and cranial size for an intraspecific 'ontogenetic' model of dwarfing.** Major axis slopes (95% confidence intervals): thin line, *H. amphibiaus*, 0.468 (0.44–0.50); bold line, dwarf species pooled, 0.454 (0.27–0.67). Filled diamonds, *H. amphibiaus* ( $n = 37$ ); filled triangles, *H. lemerlei* ( $n = 12$ ); open triangles, *H. madagascariensis* ( $n = 12$ ); crosses, *C. liberiensis* ( $n = 6$ ). For statistical comparisons see Supplementary Table 1.

especially given the magnitude of the dwarfing (Table 2). Cranial volume, determined by direct measurement<sup>22</sup>, is reduced only 50-fold, and when scaled as (cranial volume)<sup>0.35</sup>, *P. falconeri* would have an endocranial capacity (2,250 cm<sup>3</sup>) 25% of the original value of its ancestor (*P. antiquus*) 50 times its size (Tables 1 and 2). The actual endocranial capacity of *P. falconeri* is 1,800 cm<sup>3</sup> (ref. 22), 20% smaller than the value predicted from this 'late ontogenetic' model (Table 1, first entry for *P. falconeri*), in line with our results on the dwarf hippos. This discrepancy between the predictions based on cranial size<sup>22</sup> and those based on estimates of body mass<sup>23,25</sup> in these elephants gives grounds for caution in the use of the latter for the interpretation of brain-size scaling (see refs 7, 10, 11).

One argument put forward to rebut the idea that the small brain of *H. floresiensis* was linked to the process of insular dwarfism was that its small brain size could not be accommodated within predictions made from mammalian intraspecific brain–body scaling models<sup>10,11</sup>. Here we show empirically that it is mechanistically possible for dwarf mammals on islands to evolve significantly smaller brains in relation to their cranial size than would be predicted from models of dwarfing based on intraspecific scaling of the mainland ancestor. If the hippo model is applied to a typical *H. erectus* ancestor with a body mass of 60 kg (refs 10, 11, 26) and average endocranial capacity of 991 cm<sup>3</sup> (refs 10, 11, 27) that reduced its body mass by 62% to 23 kg (the median of the 16–29 kg estimated body mass of *H. floresiensis*)<sup>6</sup>, an endocranial capacity of 704 cm<sup>3</sup> (71% of the original value) would be predicted from scaling as (body mass)<sup>0.35</sup>. If the brain size were decreased by a further 30% of that value, as in the Malagasy hippo, *H. lemerlei*, an endocranial capacity of 493 cm<sup>3</sup> would result. This is still larger than the actual value of *H. floresiensis* (380–430 cm<sup>3</sup>)<sup>6,28</sup>, but if the ancestor had an endocranial capacity of 804 cm<sup>3</sup>, as in the African *H. erectus* KNM-ER 3883 (ref. 27), scaling the body mass of the ancestor down to 23 kg and then decreasing the brain size by a further 30% of the scaled value gives an endocranial capacity of 405 cm<sup>3</sup>, comparable to that of *H. floresiensis* (see Supplementary Table 5). By the same analogy, if the Dmanisi adult *Homo* remains (skull D3444 cf. *H. erectus*)<sup>29</sup>, with an endocranial capacity 650 cm<sup>3</sup> (ref. 29) and body mass of 40 kg (ref. 30) are considered, an endocranial capacity of 378 cm<sup>3</sup> would result, a value close to that of *H. floresiensis* (see Supplementary Table 5). If cranial variables are used instead of body mass for scaling the endocranial capacity of these hominins, as in the hippo example, a similar result is obtained (see Supplementary Table 5).

Similar shifts in intercept to those illustrated in Fig. 1 have been reported from domesticated mammals that on average have smaller brains than their wild relatives<sup>16</sup>. Domesticated mammals provide proof of the developmental malleability of the brain among closely related species or within species, but there are some noteworthy distinctions between the phenomena of domestication and island dwarfism. The decrease in brain size of domestic mammals is not necessarily associated with decrease in body size, as it is in phyletic dwarfism, and domestication is usually associated with a decrease in the size of the sense organs<sup>16</sup>. *H. floresiensis*, unlike the bovid *Myotragus* and domesticated mammals, does not have reduced orbital dimensions<sup>11</sup>, although such skeletal variation does not necessarily correlate with the actual size of the sense organ. This is also evident in the Malagasy hippos, in which orbit area is actually larger in the dwarf hippos relative to *H. amphibiaus* (Supplementary Fig. 2). A pathological explanation for this condition in the Malagasy species can be ruled out through the existence of more than 40 individual dwarf hippo specimens with intact braincases (Supplementary Table 2). Whatever the explanation for the tiny brain of *H. floresiensis* relative to its body size, the evidence presented here suggests that the phenomenon of insular dwarfism could have played a part in its evolution.

## METHODS SUMMARY

A cross-sectional postnatal series of crania of *H. lemerlei*, *H. madagascariensis*, *H. amphibiaus* and *C. liberiensis* (see Supplementary Tables 2 and 3) was digitized with a MicroScribe G2 and inter-landmark distances were calculated to compute a volumetric measure of the entire cranium. Endocranial capacity was used as a surrogate for brain size. Exponents derived from brain–body allometric scaling relationships in mammals are widely applied in analyses of relative brain size<sup>5</sup>, but here brain size has been regressed against cranial volume. Supplementary data are provided demonstrating that results generated from estimates of body mass are similar to those reported for cranial size (see Supplementary Figs 3–7 and Supplementary Discussion). The data and procedure used to determine body mass estimates from Malagasy dwarf postcrania are given in Supplementary Table 4 and in the Supplementary Discussion.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 15 January 2008; accepted 19 February 2009.**

1. Shea, B. T. Phyletic size change and brain/body allometry: a consideration based on the African pongids and other primates. *Int. J. Primatol.* **4**, 33–62 (1983).
2. Martin, R. D. *Human Brain Evolution in an Ecological Context: 52nd James Arthur Lecture on the Evolution of the Human Brain* (American Museum of Natural History, New York, 1983).
3. Gould, S. J. Allometry in primates with emphasis on scaling and the evolution of the brain. *Contrib. Primatol.* **5**, 244–292 (1975).
4. Köhler, M. & Moyà-Solà, S. Reduction of brain and sense organs in the fossil insular bovid *Myotragus*. *Brain Behav. Evol.* **63**, 125–140 (2004).
5. Martin, R. D. & Harvey, P. H. in *Size and Scaling in Primate Biology* (ed. Jungers, W. L.) 147–173 (Plenum, 1985).
6. Brown, P. et al. A new small-bodied hominin from the Late Pleistocene of Flores, Indonesia. *Nature* **431**, 1055–1061 (2004).
7. Niven, J. E. Brains, islands and evolution: breaking all the rules. *Trends Ecol. Evol.* **22**, 57–59 (2006).
8. Richards, G. D. Genetic, physiologic and ecogeographic factors contributing to variation in *Homo sapiens*: *Homo floresiensis* reconsidered. *J. Evol. Biol.* **19**, 1744–1767 (2006).
9. Lalueza-Fox, C., Shapiro, B., Bover, P., Alcover, J. A. & Bertranpetit, J. Molecular phylogeny and evolution of the extinct bovid *Myotragus balearicus*. *Mol. Phylogenet. Evol.* **25**, 501–510 (2002).
10. Martin, R. D. et al. Comment on 'The brain of LB1, *Homo floresiensis*'. *Science* **312**, 999b (2006).
11. Martin, R. D., Maclarnon, A. M., Phillips, J. L. & Dobyns, W. B. Flores hominid: new species or microcephalic dwarf? *Anat. Rec.* **288A**, 1123–1145 (2006).
12. Köhler, M., Moyà-Solà, S. & Wrangham, R. W. Island rules cannot be broken. *Trends Ecol. Evol.* **23**, 6–7 (2008).
13. Argue, D., Donlon, D., Groves, C. & Wright, R. *Homo floresiensis*: microcephalic, pygmyoid, *Australopithecus*, or *Homo*? *J. Hum. Evol.* **51**, 360–374 (2006).
14. Niven, J. E. Response to Köhler et al. Impossible arguments about possible species? *Trends Ecol. Evol.* **23**, 8–9 (2008).
15. Lande, R. Quantitative genetic analysis of multivariate evolution, applied to brain:body size allometry. *Evolution Int. J. Org. Evolution* **33**, 402–416 (1979).



16. Kruska, D. C. T. On the evolutionary significance of encephalization in some eutherian mammals: effects of adaptive radiation, domestication and feralization. *Brain Behav. Evol.* **65**, 73–108 (2005).
17. Stuenkel, S. Taxonomy, habits, and relationships of the subfossil Madagascan hippopotami *Hippopotamus lemerlei* and *H. madagascariensis*. *J. Vertebr. Paleontol.* **9**, 241–268 (1989).
18. Boissarie, J.-R. The phylogeny and taxonomy of Hippopotamidae (Mammalia: Artiodactyla): a review based on morphology and cladistic analysis. *Zool. J. Linn. Soc.* **143**, 1–26 (2005).
19. Eltringham, S. K. *The Hippos: Natural History and Conservation* (Academic, 1999).
20. Burney, D. A. et al. A chronology for late prehistoric Madagascar. *J. Hum. Evol.* **47**, 25–63 (2004).
21. Weston, E. M. Evolution of ontogeny in the hippopotamus skull: using allometry to dissect developmental change. *Biol. J. Linn. Soc.* **80**, 625–638 (2003).
22. Accordi, F. S. & Palombo, M. R. Morfologia endocranica degli elefanti nani pleistocenici di Spinagallo (Siracusa) e comparazione con l'endocranio di *Elephas antiquus*. *Atti Accad. Naz. Lincei R.* **51**, 111–124 (1971).
23. Roth, V. L. Inferences from allometry and fossils: dwarfing of elephants on islands. *Oxf. Surv. Evol. Biol.* **8**, 259–288 (1992).
24. Shoshani, J., Kupsky, W. J. & Marchant, G. H. Elephant brain. Part I: Gross morphology, functions, comparative anatomy, and evolution. *Brain Res. Bull.* **70**, 124–157 (2006).
25. Christiansen, P. Body size in proboscideans, with notes on elephant metabolism. *Zool. J. Linn. Soc.* **140**, 524–549 (2004).
26. Kappelman, J. The evolution of body mass and relative brain size in fossil hominids. *J. Hum. Evol.* **30**, 243–276 (1996).
27. Stanyon, R., Consiglieri, S. & Morescalchi, M. A. Cranial capacity in hominid evolution. *Hum. Evol.* **8**, 205–216 (1993).
28. Jacob, T. et al. Pygmoid Australomelanesian *Homo sapiens* skeletal remains from Liang Bua, Flores: population affinities and pathological abnormalities. *Proc. Natl Acad. Sci. USA* **103**, 13421–13426 (2006).
29. Lordkipanidze, D. et al. A fourth hominin skull from Dmanisi, Georgia. *Anat. Rec.* **288A**, 1146–1157 (2006).
30. Lordkipanidze, D. et al. Postcranial evidence from early *Homo* from Dmanisi, Georgia. *Nature* **449**, 305–310 (2007).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank A. Curren, C. Lefèvre, C. Sagne, E. Gilissen, F. Renoult, H. Chatterjee, J. Ashby, M. Nowak-Kemp, M. Harman, P. Jenkins, P. Tassy, R. Sabin, R. Symonds and S. Stuenkel for facilitating access to museum collections; A. Rasoamimanana, G. Ravololonarivo, H. Andriamialison, T. Rakotondrazafy, M. Ramarolahy and S. Bourlat for permission and/or assistance with study of the subfossil material held in the University of Antananarivo and the Académie Malagache; B. Ramanivosoa, D. Gommery, C. Guérin and M. Faure for allowing the study of material at the Akiba Museum, Mahajanga, Madagascar; R. Portela Miguez for assistance with recording endocranial capacity measures from *H. amphibia* specimens in the Natural History Museum, London; A. Friday for assistance with data collection in the University Museum of Zoology, Cambridge; C. Anderung, J.-R. Boissarie, S. Walsh and V. Herridge for discussion and helpful comments; and J. Kappelman, J. Niven, D. Lieberman and A. Gordon for comments on earlier versions of this manuscript. This research was supported by the Biotechnology and Biological Sciences Research Council.

**Author Contributions** E.W. and A.L. designed the study. E.W. collected and analysed the data and drafted the paper. Both authors discussed the results and edited the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to E.W. (e.weston@nhm.ac.uk).

## METHODS

**Sample size.** The number of specimens included in different analyses is not constant as a result of missing data (MD); all relevant landmarks were not preserved on every specimen studied, and endocranial capacity could not be measured accurately on some specimens as a result of braincase damage or poor state of preservation (see Supplementary Tables 2 and 3). Sample sizes were: *H. lemerlei*,  $n = 29$  (17 with MD); *H. madagascariensis*,  $n = 19$  (7 with MD); *H. amphibi*,  $n = 50$  (13 with MD); *C. liberiensis*,  $n = 7$  (1 with MD).

**Estimation of cranial volume.** A volumetric measure of the entire cranium was computed from the product of three variables: cranial length (the posteriormost point of the nuchal crest to the mesialmost point of the first incisor socket), cranial width (zygion to zygion) and cranial height (akrokranium (the median dorsal point of the occipital region) to basion). The static adult samples used to derive values given in Table 2 include specimens from dental 'Age Group XI' and above (see Supplementary Tables 2 and 3 and Supplementary Discussion).

**Estimation of endocranial capacity.** Endocranial capacity, the volume of the endocranial cavity, was measured by pouring precision plastic (polypropylene) balls 5.5 mm in diameter into the braincase cavity (after the large foramina had been plugged with plastazote foam), and then decanting the balls into a measuring cylinder.

**Analysis of relative brain size.** Brain size in mammals scales allometrically with body size and is described by the bivariate power function

$$y = bx^k$$

or, after logarithmic transformation, by the linear equation

$$\log y = k \log x + \log b$$

where  $y$  and  $x$  are variables, and  $k$  (exponent expressing slope) and  $b$  ( $y$  intercept) are constants. Slopes were determined by major-axis and least-squares regression (see Supplementary Table 1), but major-axis values define  $k$  in Table 1. In view of the absence of associated data on body mass for a growth series of modern *H. amphibi* skeletons, an analysis of cranial size versus global skeletal size (an estimate of body mass determined from postcranial evidence) has been provided in Supplementary Material (see Supplementary Discussion). For the subfossil taxa, body mass values were estimated from crania or from unassociated skeletal elements (see Supplementary Table 4 and Supplementary Discussion). The endocranial cavity (used to represent brain size), forms part of the cranium; cranial size is therefore a conservative estimator of change that can be measured directly from both fossil and modern material, and in hippos it is a good proxy for body size (see Supplementary Discussion).

# Decision-related activity in sensory neurons reflects more than a neuron's causal effect

Hendrikje Nienborg<sup>1</sup> & Bruce G. Cumming<sup>1</sup>

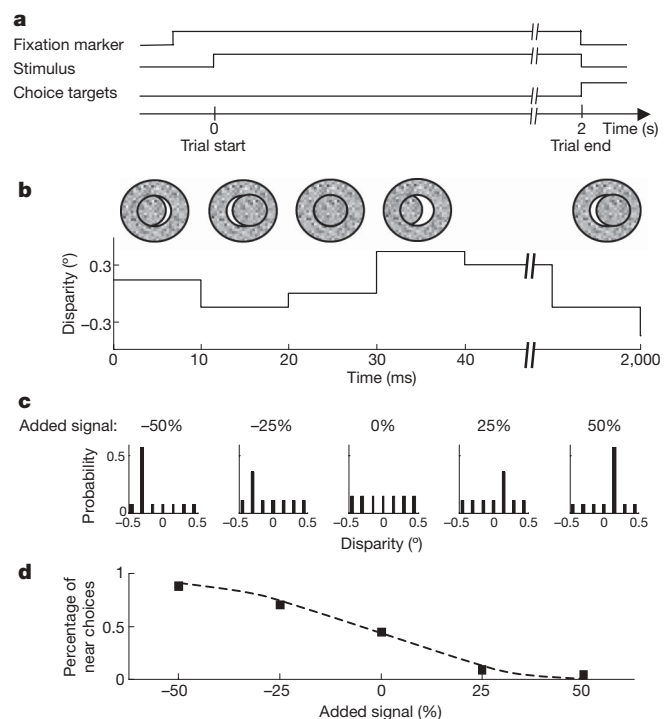
During perceptual decisions, the activity of sensory neurons correlates with a subject's percept, even when the physical stimulus is identical<sup>1–9</sup>. The origin of this correlation is unknown. Current theory proposes a causal effect of noise in sensory neurons on perceptual decisions<sup>10–12</sup>, but the correlation could result from different brain states associated with the perceptual choice<sup>13</sup> (a top-down explanation). These two schemes have very different implications for the role of sensory neurons in forming decisions<sup>14</sup>. Here we use white-noise analysis<sup>15</sup> to measure tuning functions of V2 neurons associated with choice and simultaneously measure how the variation in the stimulus affects the subjects' (two macaques) perceptual decisions<sup>16–18</sup>. In causal models, stronger effects of the stimulus upon decisions, mediated by sensory neurons, are associated with stronger choice-related activity. However, we find that over the time course of the trial these measures change in different directions—at odds with causal models. An analysis of the effect of reward size also supports this conclusion. Finally, we find that choice is associated with changes in neuronal gain that are incompatible with causal models. All three results are readily explained if choice is associated with changes in neuronal gain caused by top-down phenomena that closely resemble attention<sup>19</sup>. We conclude that top-down processes contribute to choice-related activity. Thus, even forming simple sensory decisions involves complex interactions between cognitive processes and sensory neurons.

Considerable progress has been made in explaining the neuronal mechanisms underlying decision making<sup>12</sup>, which is a major goal in systems neuroscience. For simple perceptual decisions, recent theory proposes that sensorimotor areas accumulate sensory evidence about the physical world, delivered by sensory neurons<sup>10,11,20–22</sup>. Noise in the sensory neurons causes variability in the behavioural response<sup>10–12</sup>, resulting in a covariation between the neuronal activity and the behaviour<sup>1–9</sup>. (We note that this causal effect of noise in the sensory representation has only been invoked for sensory areas, not for sensorimotor areas.) However, this covariation could also result from top-down effects<sup>13</sup> in which brain states<sup>23</sup> that are associated with one behavioural response also alter the response of the sensory neurons. A third (bottom-up) possibility is that sensory neurons that themselves have no causal effect on the decision are correlated with sensory neurons that do have a causal effect. These schemes have markedly different implications for the role of sensory neurons in forming decisions. Sensory neurons either only encode the physical stimulus or simultaneously form an integral part of the mechanism used by the brain to decode the sensory information. To distinguish between these views, we combined the measurement of choice-related activity in disparity selective V2 neurons in a disparity-discrimination task, with a stimulus that permitted the use of white-noise analysis. This allowed the simultaneous application of 'subspace mapping'<sup>15</sup>, to describe how disparity affects the neuronal response ('disparity subspace map'), and 'psychophysical reverse correlation'<sup>16–18</sup>, to extract a kernel

describing how disparity affects the subjects' perceptual choices. This comprehensive data set enables us to differentiate among these schemes.

Two macaque monkeys performed a coarse disparity-discrimination task (Fig. 1) while we recorded extracellularly from disparity-selective neurons in their visual areas V2. The stimulus, a circular random dot stereogram, contained a spatially uniform binocular disparity that varied randomly on each video frame. At the end of each trial the monkeys reported in a forced choice task whether the stimulus appeared near ('near' choice) or far ('far' choice). We exploited this random variation to perform psychophysical reverse correlation<sup>16–18</sup>, and simultaneously to measure neuronal subspace maps<sup>15</sup> for disparity.

First we examined how the monkeys weight the disparity signal in the stimulus to form their decision<sup>16</sup>. We calculated the difference between the average stimulus preceding the monkeys' 'near' choices



**Figure 1 | Methods.** **a**, Sketch of the sequence of events during one trial. **b**, Example time series of the stimulus. **c**, Probability mass distributions of the stimuli for one experiment (probability as a function of disparity), with signal disparities of  $-0.3^\circ$  and  $0.15^\circ$ . Each plot depicts one signal condition (negative percentages indicate near signal disparities). **d**, The monkey's performance for this experiment (in percentage 'near' choices as a function of percentage added signal).

<sup>1</sup>Laboratory of Sensorimotor Research, National Eye Institute, National Institutes of Health, 49 Convent Drive, Bethesda, Maryland 20892, USA.



and the average stimulus preceding the monkeys' 'far' choices. This 'psychophysical kernel' measures the relative probability with which the disparity on any given frame occurred before the monkeys' 'near' choice. The amplitude of the kernel decreased substantially over the course of the trial (Fig. 2a, b). (The Supplementary Information discusses the shape of the psychophysical kernel and shows that this linear analysis adequately captures the monkeys' behaviour.) This means that the monkeys relied predominantly on the stimulus disparities at the start of the trial and progressively less so towards its end. If one considers neurons representing this sensory evidence, their activity early in the trial should have a stronger effect on the decision than activity late in the trial. Thus, if the choice-related activity reflected only the causal effect of the neuronal firing on the choice, the size of the choice-related activity should also decrease over time. This prediction follows directly from the fact that, in the causal explanation, choice-related activity is the effect of noise in the sensory evidence that is used to make a decision.

To test this prediction, we quantified the choice-related signal as 'choice probability'<sup>3</sup>. (Choice probabilities were corrected for the stimulus-induced component; see Supplementary Information.) The time course of the choice-related signal in our data (Fig. 2c) is different from that predicted from the time course of the psychophysical data in the causal-only scheme. Choice probabilities were measured for 76 neurons that had been recorded while the data for the psychophysical

kernel were gathered. For 57 of the 76 neurons, for which the choice probability was  $>0.5$ , we examined the mean choice probability as a function of time (Fig. 2c). Consistent with previous findings<sup>4</sup>, the choice probability plateaus after about 500 ms, which is quite unlike the statistically significant decrease in amplitude of the psychophysical kernel over time (correlation coefficient,  $r = -0.81$ ;  $P < 10^{-23}$  between amplitude and time, over the second half of the trials). Although choice-probability time courses for individual neurons are noisy, we addressed the possibility that some neurons behave as if they play a causal role by computing the correlation coefficient,  $R$ , between the time course of the choice probability for each individual neuron and the time course of the average psychophysical kernel amplitude (Fig. 2d). We found a significant negative correlation between these coefficients and a neuron's choice probability ( $r = -0.28$ ,  $P < 0.05$ ), indicating that neurons with high choice probabilities tended to show a negative correlation with the time course of the psychophysical kernel amplitude, as expected from the average data (Fig. 2b). This and other features of individual time courses (Supplementary Discussion) are at odds with the causal model.

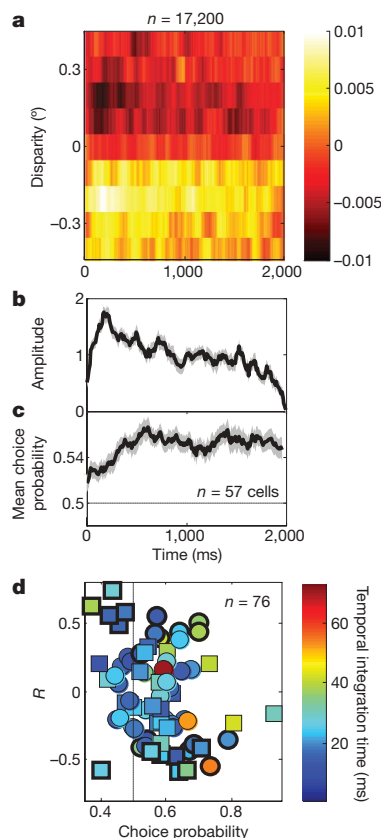
The fact that the results in Fig. 2 are incompatible with the causal-only account, suggests that choice probabilities are at least partly of non-causal, possibly top-down, origin. We therefore sought a signature of possible top-down mechanisms at the level of individual neurons. This could employ a mechanism, similar to attention, that characteristically alters the gain of sensory neurons<sup>19</sup>. To test this possibility, we designed our disparity-varying stimulus such that it permitted the measurement of subspace maps for disparity (Methods).

These subspace maps quantify the effect of each disparity (in the stimulus with no added signal) on the neuron. Calculating subspace maps separately for stimuli associated with 'near' and 'far' choices quantifies any effects of choice on the neuronal response. Choice-related activity itself implies some difference between these subspace maps. If the difference is caused by a change in neuronal gain, the two subspace maps should be scaled versions of each other. Example subspace maps for one neuron (Fig. 3a, b) show that the gain of this neuron increased by 84%, whereas the additive change was close to 0 ( $-0.032$  spikes per frame). A second example shows a more typical gain increase (18%;  $y$  offset, 0.005 spikes per frame; Fig. 3c, d).

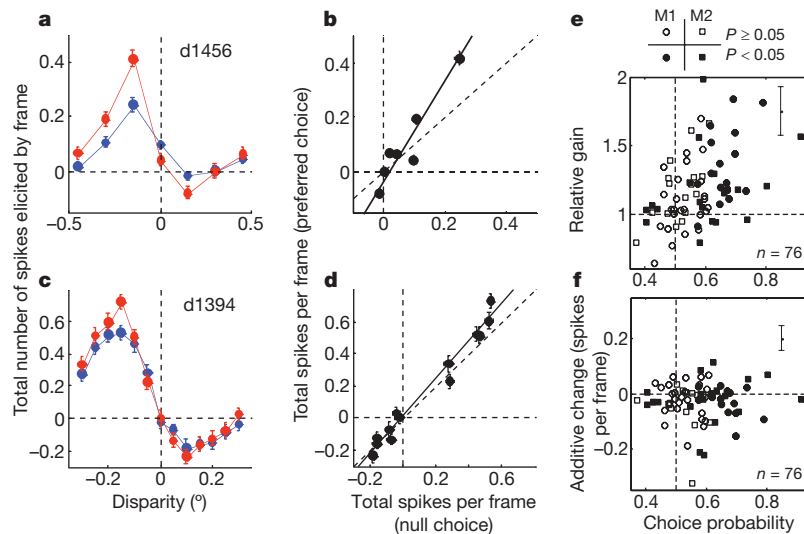
The distribution of relative gain change as a function of choice probability demonstrates that choice probabilities are associated with choice-related changes in neuronal gain ( $n = 76$ ,  $r = 0.44$ ,  $P < 10^{-4}$  (monkey 1:  $n = 42$ ,  $r = 0.54$ ,  $P < 0.001$ ; monkey 2:  $n = 34$ ,  $r = 0.32$ ,  $P < 0.07$ ); Fig. 3e). The geometric mean of the relative gains was 1.16 (1.17 and 1.15 for monkeys 1 and 2, respectively), which is significantly different from 1 ( $P < 0.001$ , by resampling). Conversely, there was no systematic relationship between the  $y$  offset and the choice probability ( $r = 0.03$ ,  $P = 0.77$  (monkey 1:  $r = -0.18$ ,  $P = 0.25$ ; monkey 2:  $r = 0.18$ ,  $P = 0.31$ ); mean offset,  $-0.03$  spikes per frame; Fig. 3f). Thus, it is the choice-dependent change in response gain that explains the difference in mean response rate between preferred and null choices. A modest gain change could arise, even in the causal account of choice probability, from the firing properties of cortical neurons (for example Poisson spiking<sup>24</sup>). A shuffling technique showed that this effect was too small to account for the observed gain changes (Supplementary Information).

The gain change suggests the operation of a mechanism similar to feature-selective attention<sup>19</sup>, but which varies from trial to trial. This could arise in several ways. First, as the decision is formed, a signal altering the neuronal gain may be sent back to those neurons supporting this decision. Alternatively, this gain change may implement a perceptual working memory<sup>25</sup>, or a perceptual bias/expectation: attending to near features increases the response gain of near-prefering neurons and thus makes a 'near' response more likely.

An additional feature of our data provides evidence that at least the latter mechanism operates. The reward size depended systematically on the animals' performance (Methods). This performance was better in trials for which a large reward was available (Fig. 4a and



**Figure 2 | Psychophysical kernel and choice-related signal have different time courses.** **a**, Psychophysical kernel (averaged over 76 experiments;  $n = 17,200$  trials; two monkeys) as a function of disparity and time. Colour represents amplitude (in occurrences per frame). **b**, Normalized amplitude of the psychophysical kernels decreases over time. **c**, Averaged choice-related signal over time. Shaded grey areas in **b** and **c**,  $\pm 1$  standard error. **d**, The correlation coefficient,  $R$ , over time between choice probability (for individual neurons) and the amplitude of the mean psychophysical kernel, plotted against a neuron's mean choice probability. Colour represents temporal integration time (Supplementary Methods); bold symbols, significant  $R$  ( $P < 0.05$ , by resampling); circles, data from monkey 1; squares, data from monkey 2.



**Figure 3 | Choice-dependent gain change.** **a**, Subspace maps for preferred (red) and null (blue) choices superimposed (neuron d1456). Dashed lines: 0° disparity, 0 spikes per frame. **b**, Null-choice responses plotted against preferred-choice responses, yielding relative gain (slope, 1.84) and additive change ( $y$  offset,  $-0.032$  spikes per frame). Dashed lines: unit relative gain, 0 spikes per frame. **c**, **d**, Same format as in **a** and **b** for neuron d1394, the slope (1.18) and  $y$  offset (0.005 spikes per frame) of which resemble the population

mean. Error bars in **a–d**, standard errors (by resampling). **e**, Slope and choice probability are correlated. Filled and open symbols respectively represent cells with and without significant choice probability. Circles, data for monkey 1 (M1); squares, data for monkey 2 (M2). Dashed lines: 0.5 choice probability, unit relative gain. **f**, There is no correlation between  $y$  offset and choice probability (symbols as in **e**). Dashed lines: 0.5 choice probability, 0 spikes per frame. Error bars in **e** and **f**, median standard error.

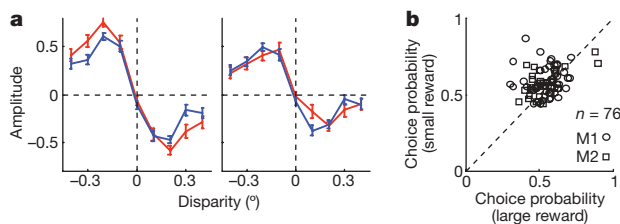
Supplementary Information), indicating that animals used more information about the stimulus when reward size was large. It allows us to further test causal explanations for choice probability: when the animal uses more stimulus-derived information, choice probability should be larger. Contrary to this prediction, we found that choice probabilities were significantly smaller for trials in which a large reward was available ( $P < 0.006$ , paired  $t$ -test; Fig. 4b). This result can be explained by assuming that the animal has some bias (or expectation) at the start of each trial (regardless of reward size), and that this bias engages our proposed top-down mechanism. When the available reward is small and the monkeys make less use of the sensory input (as demonstrated by the psychophysical kernel; Fig. 4a), the bias will have a stronger impact on the behavioural response. Conversely, when a large reward is available, the improved performance implies that any initial bias is more likely to be overridden by the evidence provided by the visual stimulus. Hence, any component of choice probability reflecting a top-down effect of bias will be smaller in large-reward trials when the decision is more strongly driven by the actual stimulus and less by the monkey's initial bias.

Our results provide three lines of evidence against the view that decision-related activity in sensory neurons reflects only the causal effect of neuronal noise on sensory decisions. First, the time course of the decision-related signal was incompatible with that predicted from the behavioural data in the causal-only scheme. Second, larger rewards systematically improved the animals' behaviour, but reduced choice probability, a result opposite to the expectation from causal explanations. Finally, choice probabilities were associated with choice-dependent gain changes larger than could be explained in the causal scheme. All three phenomena follow naturally from a top-down scheme in which the animals' perceptual state alters the response of sensory neurons. An alternative explanation is that neurons which do not contribute to the decision show choice probability because they are correlated with neurons that do contribute. Such a scheme, if sufficiently rich, may explain the data without invoking a top-down mechanism (Supplementary Discussion), but nonetheless abandons the principle that choice probabilities reflect only the causal effect of sensory noise upon decisions. Given that the choice-dependent gain changes we observe are characteristic of top-down mechanisms such as attention, our top-down scheme is the most parsimonious.

Changes in neuronal gain may facilitate the decoding of neuronal populations by appropriately weighting relevant neurons<sup>26–28</sup>. Implementing such a decoding mechanism at the level of sensory neurons allows the brain extraordinary flexibility to perform sensory decisions in different circumstances. Here we have shown that these gains vary with a subject's choice, within a fixed task. This gain change could implement a perceptual bias or expectation (attending to near or far features), and could also follow the formation of a decision. It may serve to promote perceptual stability in the presence of ambiguous<sup>29</sup> or noisy sensory signals. Whether it reflects an expectation or a post-decision signal, our data suggest that even simple sensory decisions involve top-down mechanisms that combine cognitive processes and the sensory representation in previously unreported ways.

## METHODS SUMMARY

All procedures were performed in accord with the US Public Health Service policy on the humane care and use of laboratory animals and all protocols were approved by the National Eye Institute Animal Care and Use Committee. We recorded extracellular activity from disparity-selective single V2 units while two



**Figure 4 | Reward size affects behaviour and choice-related signal.** **a**, Psychophysical kernel as a function of disparity and available reward, in occurrences per 1,000 ms, averaged over the first (left panel) and second (right panel) 1,000 ms of each trial ( $n = 6,886$  trials for large reward (red);  $n = 10,314$  trials for small reward (blue)). Improved performance is mainly caused by there being a larger psychophysical kernel in the first 1,000 ms of the trials (kernel difference  $P < 0.001$ , by resampling), but not the second half (difference not significant). Error bars, standard errors (by resampling). **b**, Choice probability computed for the first 1,000 ms of the trials was larger when a smaller reward was available ( $P < 0.006$ ,  $n = 76$ ). Dashed line: unit slope.

monkeys (*Macaca mulatta*) performed disparity discrimination. Upon fixation, a stimulus was presented for 2 s and followed by two choice targets. After a saccade to the correct target, the monkeys received a liquid reward. Stimuli were dynamic random dot patterns: a disparity-varying centre (disparity changed randomly on each frame; 96-Hz frame rate) and a surrounding annulus at 0°. The centre disparity was chosen from a set of evenly spaced disparity values centred around 0° (encompassing the preferred and null disparity of each neuron). We introduced a detectable signal by increasing the probability of occurrence of one disparity within some trials. These signal disparities approximately matched each neuron's preferred and null disparity. Signal trials served only to control behaviour: all analyses were restricted to trials with no signal added.

Psychophysical kernels were computed as the difference in the respective mean stimulus matrices before 'near' and 'far' choices. The average kernel was a weighted average of the kernel for each experiment for which neuronal data were included. Choice probabilities were obtained as described previously<sup>3,6</sup>, but corrected for the stimulus-induced component (Supplementary Information). For the subspace analysis, the average response of each neuron following one frame of a given disparity,  $d_i$ , was computed as a spike density function,  $S_i(t)$ . We calculated the total number of spikes elicited by one frame of disparity  $d_i$  as the sum of the overall mean number of spikes per frame and the integral of the deviation of  $S_i(t)$  about this mean. Analysing the trials separately according to the monkeys' choice yield subspace maps for each choice.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 29 October 2008; accepted 23 January 2009.

Published online 8 March 2009.

1. Logothetis, N. K. & Schall, J. D. Neuronal correlates of subjective visual perception. *Science* **245**, 761–763 (1989).
2. Leopold, D. A. & Logothetis, N. K. Activity changes in early visual cortex reflect monkeys' percepts during binocular rivalry. *Nature* **379**, 549–553 (1996).
3. Britten, K. H., Newsome, W. T., Shadlen, M. N., Celebrini, S. & Movshon, J. A. A relationship between behavioral choice and the visual responses of neurons in macaque MT. *Vis. Neurosci.* **13**, 87–100 (1996).
4. Uka, T. & DeAngelis, G. C. Contribution of area MT to stereoscopic depth perception: choice-related response modulations reflect task strategy. *Neuron* **42**, 297–310 (2004).
5. Uka, T., Tanabe, S., Watanabe, M. & Fujita, I. Neural correlates of fine depth discrimination in monkey inferior temporal cortex. *J. Neurosci.* **25**, 10796–10802 (2005).
6. Nienborg, H. & Cumming, B. G. Macaque V2 neurons, but not V1 neurons, show choice-related activity. *J. Neurosci.* **26**, 9567–9578 (2006).
7. Dodd, J. V., Krug, K., Cumming, B. G. & Parker, A. J. Perceptually bistable three-dimensional figures evoke high choice probabilities in cortical area MT. *J. Neurosci.* **21**, 4809–4821 (2001).
8. Purushothaman, G. & Bradley, D. C. Neural population code for fine perceptual decisions in area MT. *Nature Neurosci.* **8**, 99–106 (2005).
9. Ress, D. & Heeger, D. J. Neuronal correlates of perception in early visual cortex. *Nature Neurosci.* **6**, 414–420 (2003).
10. Shadlen, M. N., Britten, K. H., Newsome, W. T. & Movshon, J. A. A computational analysis of the relationship between neuronal and behavioral responses to visual motion. *J. Neurosci.* **16**, 1486–1510 (1996).
11. Yang, T. & Shadlen, M. N. Probabilistic reasoning by neurons. *Nature* **447**, 1075–1080 (2007).
12. Schall, J. D. Neural correlates of decision processes: neural and mental chronometry. *Curr. Opin. Neurobiol.* **13**, 182–186 (2003).
13. Krug, K. A common neuronal code for perceptual processes in visual cortex? Comparing choice and attentional correlates in V5/MT. *Phil. Trans. R. Soc. Lond. B* **359**, 929–941 (2004).
14. Gilbert, C. D. & Sigman, M. Brain states: top-down influences in sensory processing. *Neuron* **54**, 677–696 (2007).
15. Ringach, D. L., Hawken, M. J. & Shapley, R. Dynamics of orientation tuning in macaque primary visual cortex. *Nature* **387**, 281–284 (1997).
16. Neri, P., Parker, A. J. & Blakemore, C. Probing the human stereoscopic system with reverse correlation. *Nature* **401**, 695–698 (1999).
17. Ahumada, A. J. Perceptual classification images from Vernier acuity masked by noise. *Perception* **26**, 18 (1996).
18. Nienborg, H. & Cumming, B. G. Psychophysically measured task strategy for disparity discrimination is reflected in V2 neurons. *Nature Neurosci.* **10**, 1608–1614 (2007).
19. Treue, S. & Martinez Trujillo, J. C. Feature-based attention influences motion processing gain in macaque visual cortex. *Nature* **399**, 575–579 (1999).
20. Gold, J. I. & Shadlen, M. N. Representation of a perceptual decision in developing oculomotor commands. *Nature* **404**, 390–394 (2000).
21. Gold, J. I. & Shadlen, M. N. The neural basis of decision making. *Annu. Rev. Neurosci.* **30**, 535–574 (2007).
22. Mazurek, M. E., Roitman, J. D., Ditterich, J. & Shadlen, M. N. A role for neural integrators in perceptual decision making. *Cereb. Cortex* **13**, 1257–1269 (2003).
23. Arieli, A., Sterkin, A., Grinvald, A. & Aertsen, A. Dynamics of ongoing activity: explanation of the large variability in evoked cortical responses. *Science* **273**, 1868–1871 (1996).
24. Dean, A. F. The variability of discharge of simple cells in the cat striate cortex. *Exp. Brain Res.* **44**, 437–440 (1981).
25. Pasternak, T. & Greenlee, M. W. Working memory in primate sensory systems. *Nature Rev. Neurosci.* **6**, 97–107 (2005).
26. Ma, W. J., Beck, J. M., Latham, P. E. & Pouget, A. Bayesian inference with probabilistic population codes. *Nature Neurosci.* **9**, 1432–1438 (2006).
27. Martinez-Trujillo, J. C. & Treue, S. Feature-based attention increases the selectivity of population responses in primate visual cortex. *Curr. Biol.* **14**, 744–751 (2004).
28. Jazayeri, M. & Movshon, J. A. Optimal representation of sensory information by neural populations. *Nature Neurosci.* **9**, 690–696 (2006).
29. Leopold, D. A., Wilke, M., Maier, A. & Logothetis, N. K. Stable perception of visually ambiguous patterns. *Nature Neurosci.* **5**, 605–609 (2002).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This research was supported by the Intramural Research Program of the US National Institutes of Health, National Eye Institute. We are grateful to J. A. Movshon and M. Shadlen for discussions and to the members of the Laboratory of Sensorimotor Research for comments on an earlier version of this manuscript. We also thank D. Parker and B. Nagy for excellent animal care.

**Author Contributions** H.N. designed the project, collected the data, performed the analyses and wrote the paper. B.G.C. supervised the project.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to H.N. ([hniemb@gmail.com](mailto:hniemb@gmail.com)).



## METHODS

**Task-and-reward regimen.** Two monkeys were trained in a binary, forced-choice disparity-discrimination task (Fig. 1a). They judged whether the central stimulus region appeared in front or behind the surrounding annulus. Trials started upon fixation (within  $0.5^\circ$  of a fixation marker), initiating a 2-s stimulus presentation followed by the appearance of two choice targets (respectively  $3^\circ$  above and below the fixation marker). If the monkeys indicated their decision by a saccade to the correct choice target within 500 ms of the stimulus disappearance, they received liquid rewards. If the monkeys made correct choices on three consecutive trials, the reward on the fourth and on all subsequent correct trials was approximately three times its normal size, until the monkey made an error. After an error, the reward size was reset to its normal size.

**Recordings.** We recorded extracellular activity from disparity-selective single units in these monkeys' visual areas V2, as described previously<sup>6,30</sup>. Both animals were implanted with scleral search coils in both eyes<sup>31</sup>, head-fixation posts and a recording chamber under general anaesthesia. Positions of both eyes (for 17 of 58 neurons for monkey 2, signals were available only for one eye) were measured with a magnetic scleral search system (CNC Engineering) and digitized at 800 Hz. The monkeys viewed the stimuli on EIZO Flexscan F980 monitors in a Wheatstone stereoscope configuration (89-cm viewing distance). All procedures were performed in accord with the US Public Health Service policy on the humane care and use of laboratory animals and all protocols were approved by the National Eye Institute Animal Care and Use Committee.

**Stimulus.** All stimuli were circular, dynamic, random dot stereograms (50% black and 50% white dots of 99% contrast; dot density, generally 40%; dot size,  $0.09^\circ \times 0.09^\circ$ ). Each random dot stereogram had a disparity-varying centre ( $3\text{--}5^\circ$  in diameter) and a  $1\text{--}2^\circ$ -wide surrounding annulus at  $0^\circ$  disparity (Fig. 1b). On each frame, all centre dots had the same disparity, but this disparity value changed randomly from frame to frame (96-Hz frame rate). For the condition with no added signal, the disparity on each frame was drawn at random from a uniform distribution of discrete, equally spaced disparities (symmetrical about  $0^\circ$  disparity (central plot in Fig. 1c), encompassing the preferred and the null disparity of each neuron). Signal disparities (always one near and one far disparity) approximately matched the preferred and null disparities of the neuron. Disparity signal was introduced by increasing the probability of the signal disparity on each frame (Fig. 1c).

**Psychophysical reverse correlation.** Only trials with no added signal were included in the analysis. Each stimulus trial was summarized in a two-dimensional matrix in which each row corresponds to one disparity and each column to one stimulus frame. For each column in this matrix, there is one entry that is 1, corresponding to the disparity on this frame, and all other entries are 0. We then computed the average matrices preceding the monkey's 'near' and 'far' choices. For each of the 200 stimulus frames, the resulting values correspond to the probabilities with which each disparity preceded a 'near' choice or a 'far' choice, respectively. This yields a two-dimensional (time versus disparity) probability distribution. The difference between the probability distributions preceding 'near' choices and 'far' choices defined the psychophysical kernel for each experiment. (Negative disparities were defined as near.)

The kernel shapes change little between monkeys or as a function of the signal disparities (Supplementary Fig. 2). We therefore collapsed all the data into a single psychophysical kernel to maximize the temporal resolution. The average psychophysical kernel (Fig. 2a) was obtained for all experiments for which the simultaneously recorded neurons passed the inclusion criteria. Because the disparity range was adjusted for each neuron, the psychophysical kernel for each experiment was weighted by the number of disparity values included in this experiment (this ranged between 5 and 13 disparity values) and by the number of trials for this experiment. Only data for disparity values ( $-0.4^\circ$ ,  $-0.3^\circ$ , ...,  $0.4^\circ$ ) were included in this average.

As an estimate of the amplitude of the psychophysical kernel, we computed the inner product of the time-averaged psychophysical kernel and the psychophysical kernel (temporally smoothed, 93-ms boxcar window) for each 10-ms bin, and normalized this inner product by its overall mean. Confidence intervals for all measures were derived by resampling.

All analyses were based on the linear kernel of the psychophysical data. Consistency analyses (Supplementary Information) showed that this linear kernel provides an excellent description of the monkeys' behaviour. Further analyses indicated that second-order interactions were negligible (Supplementary Information).

**Subspace analysis.** The analysis was based on all trials with no added signal. First, the average response of each neuron following one frame of a given disparity,  $d_i$ , was computed as a spike density function,  $S_i(t)$ , smoothed using a 10-ms boxcar window (coloured solid lines in Supplementary Fig. 3a). As an estimate of the impact of one frame of disparity  $d_i$  on the firing rate of the neuron, we calculated the total number of spikes,  $s_i$ , elicited by one frame of  $d_i$ . This metric corresponds to the sum of the mean number of spikes per frame,  $\mu$  (black line in Supplementary Fig. 3a) and the integral of the deviation of  $S_i(t)$  around this mean:

$$s_i = \mu + \int (S_i(t) - \mu) dt$$

We plot the disparity subspace map,  $s_i$ , as a function of  $d_i$  in Supplementary Fig. 3b. Performing this analysis on the trials separated according to the monkeys' choice yields the subspace maps separated by choice (Fig. 3a, c). To quantify the choice-dependent modulation in tuning, we plotted the responses in the null-choice trials against those in the preferred-choice trials (Fig. 3b, d), and estimated (using type II regression) the slope (gain change) and the  $y$  offset (additive change). We note that because the spike density function is a mean rate calculated separately for each choice, variations in the disparity content of the stimulus that are associated with choice will not produce differences in the subspace maps.

**Analysis of choice probabilities.** Choice probabilities were obtained for all trials with no added signal as described previously<sup>3,6</sup>. As the psychophysical kernel demonstrates, there are systematic differences in the stimuli preceding the monkeys' choices. Choice probabilities were corrected for this stimulus-induced component (Supplementary Methods).

30. Cumming, B. G. & Parker, A. J. Binocular neurons in V1 of awake monkeys are selective for absolute, not relative, disparity. *J. Neurosci.* **19**, 5602–5618 (1999).
31. Judge, S. J., Richmond, B. J. & Chu, F. C. Implantation of magnetic search coils for measurement of eye position: an improved method. *Vision Res.* **20**, 535–538 (1980).

# Compound vesicle fusion increases quantal size and potentiates synaptic transmission

Liming He<sup>1\*</sup>, Lei Xue<sup>1\*</sup>, Jianhua Xu<sup>1</sup>, Benjamin D. McNeil<sup>1</sup>, Li Bai<sup>1</sup>, Ernestina Melicoff<sup>2</sup>, Roberto Adachi<sup>2</sup> & Ling-Gang Wu<sup>1</sup>

Exocytosis at synapses involves fusion between vesicles and the plasma membrane<sup>1</sup>. Although compound fusion between vesicles<sup>2,3</sup> was proposed to occur at ribbon-type synapses<sup>4,5</sup>, whether it exists, how it is mediated, and what role it plays at conventional synapses remain unclear. Here we report the existence of compound fusion, its underlying mechanism, and its role at a nerve terminal containing conventional active zones in rats and mice. We found that high potassium application and high frequency firing induced giant capacitance up-steps, reflecting exocytosis of vesicles larger than regular ones, followed by giant down-steps, reflecting bulk endocytosis. These intense stimuli also induced giant vesicle-like structures, as observed with electron microscopy, and giant miniature excitatory postsynaptic currents (mEPSCs), reflecting more transmitter release. Calcium and its sensor for vesicle fusion, synaptotagmin, were required for these giant events. After high frequency firing, calcium/synaptotagmin-dependent mEPSC size increase was paralleled by calcium/synaptotagmin-dependent post-tetanic potentiation. These results suggest a new route of exocytosis and endocytosis composed of three steps. First, calcium/synaptotagmin mediates compound fusion between vesicles. Second, exocytosis of compound vesicles increases quantal size, which increases synaptic strength and contributes to the generation of post-tetanic potentiation. Third, exocytosed compound vesicles are retrieved via bulk endocytosis. We suggest that this vesicle cycling route be included in models of synapses in which only vesicle fusion with the plasma membrane is considered<sup>1</sup>.

We performed cell-attached capacitance recordings at the release face of calyces in brainstem (Fig. 1a)<sup>6</sup>. Brainstem slices were from rats, unless otherwise mentioned. Before KCl application (control), spontaneous up-steps occurred at  $0.008 \pm 0.001$  Hz ( $n = 78$  patches, Fig. 1a, see Methods). About 94% of up-steps were  $<220$  aF (mean =  $84 \pm 4$  aF,  $n = 96$  up-steps). The remaining 6% were  $220$ – $380$  aF (mean =  $275 \pm 22$  aF,  $n = 6$  up-steps, Fig. 1b). A regular vesicle's capacitance is  $\sim 23$ – $220$  aF, because the vesicle diameter is  $\sim 30$ – $80$  nm in calyces<sup>7</sup>, and the specific membrane capacitance is  $8$ – $11$  fF  $\mu\text{m}^{-2}$  (ref. 8). Thus, most up-steps in the control were caused by single vesicle fusion (see also ref. 6).

KCl application ( $\sim 50$ – $100$  mM, unless mentioned) increased the up-step frequency to  $0.16 \pm 0.02$  Hz ( $n = 17$  patches, Fig. 1a, see Methods). The up-step amplitude distribution and cumulative probability curve were shifted to the right, compared to control (Fig. 1b, Kolmogorov–Smirnov test,  $P < 0.001$ ). The mean of all up-steps was  $193 \pm 9$  aF ( $n = 857$ ), or  $\sim 203\%$  of control ( $95 \pm 5$  aF,  $n = 102$ ,  $P < 0.01$ ). About 20% of up-steps were  $> 220$  aF, a regular vesicle's maximal capacitance (Fig. 1a, b), and were called 'giant' up-steps. Their mean was  $571 \pm 37$  aF ( $n = 174$  up-steps) and the largest

was  $2,878$  aF, or  $34$ – $40$  times the mean of regular vesicles ( $73$  aF in ref. 6,  $84$  aF here). Both the percentage and the mean size of giant up-steps were significantly higher than control ( $\chi^2$  test and  $t$ -test,  $P < 0.01$ ). Even when up-steps between  $100$  and  $220$  aF were measured, KCl significantly increased the amplitude from  $133 \pm 6$  aF ( $n = 28$ ) in control to  $144 \pm 2$  aF ( $n = 294$ ,  $P < 0.05$ ), suggesting compound fusion between  $2$ – $3$  regular vesicles. Thus, up-step size increase was not limited to giant up-steps.

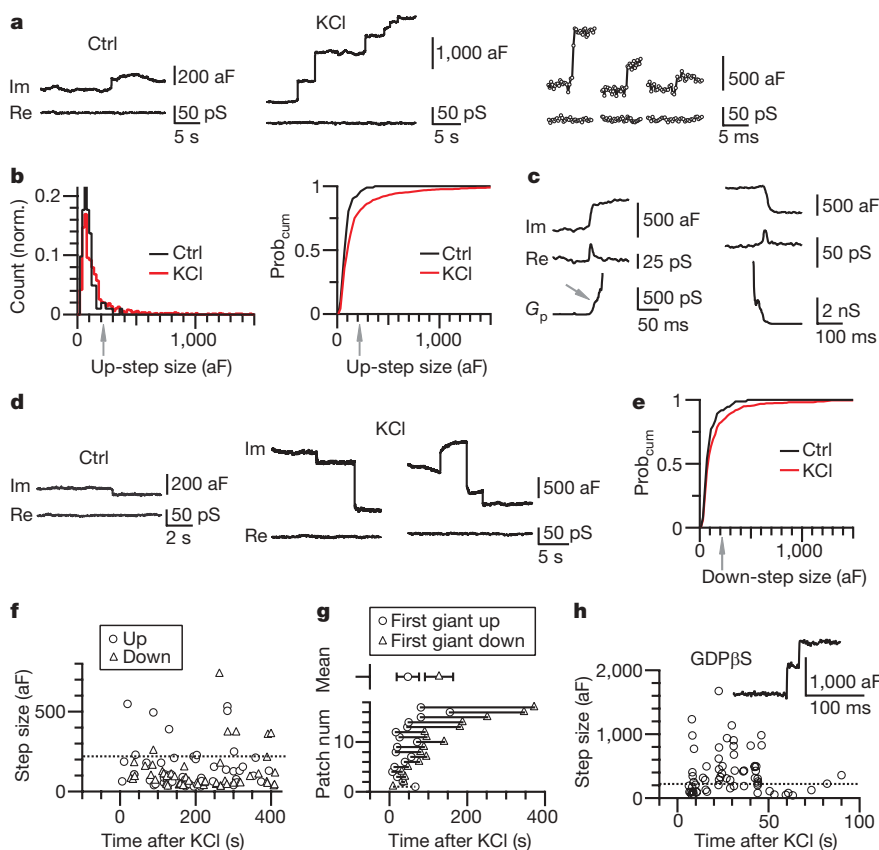
Giant up-steps were not caused by dense-core vesicle fusion, as dense-core vesicles are rare in calyces<sup>7</sup>. Giant up-steps were not composed of multiple small up-steps, because, even when the time resolution was increased to  $0.6$ – $0.9$  ms, giant up-steps rose from  $20\%$  to  $80\%$  in one step ( $n = 44$  giant up-steps, 3 patches, Fig. 1a). In 6% of giant up-steps, we detected an initial fusion pore conductance ( $G_p$ ) of  $162 \pm 41$  pS ( $n = 10$  up-steps), which gradually increased for  $\sim 10$ – $230$  ms, and ended in rapid pore expansion (Fig. 1c).  $G_p$  in remaining giant up-steps was too large or too fast to resolve (Fig. 1a). Most giant up-steps could not be explained by independent fusion of multiple vesicles, because the up-step frequency ( $\sim 0.16$  Hz) was too low to account for the observed large percentage ( $\sim 20\%$ ) of giant up-steps (Supplementary Information II-1). Thus, giant up-steps reflect fusion of a single structure larger than regular vesicles.

KCl application also increased the frequency of down-steps, which reflected endocytosis, from  $0.006 \pm 0.002$  Hz in control ( $n = 78$  patches) to  $0.10 \pm 0.01$  Hz ( $n = 17$  patches;  $P < 0.01$ , Fig. 1d), and shifted the down-step size distribution (Supplementary Information II-2) and cumulative probability curve to the right (Fig. 1e, Kolmogorov–Smirnov test,  $P < 0.01$ ). The percentage ( $17\%$ ) and the mean amplitude ( $501 \pm 42$  aF,  $n = 102$  down-steps; range:  $220$ – $2,943$  aF) of giant down-steps during KCl application were significantly higher than control ( $8\%$ ,  $\chi^2$  test,  $P < 0.01$ ;  $331 \pm 31$  aF,  $n = 6$ ,  $P < 0.01$ ). In 5% of giant down-steps, we detected an initial fission pore conductance of  $376 \pm 113$  pS ( $n = 5$ ), corresponding to a pore with a diameter of  $2.7 \pm 1.5$  nm, which reduced at  $36 \pm 11$  pS  $\text{ms}^{-1}$  or  $0.82 \pm 0.46$  nm  $\text{ms}^{-1}$  to an undetectable level (Fig. 1c). Thus, KCl induced more and larger giant down-steps, which reflect bulk endocytosis<sup>9–11</sup>.

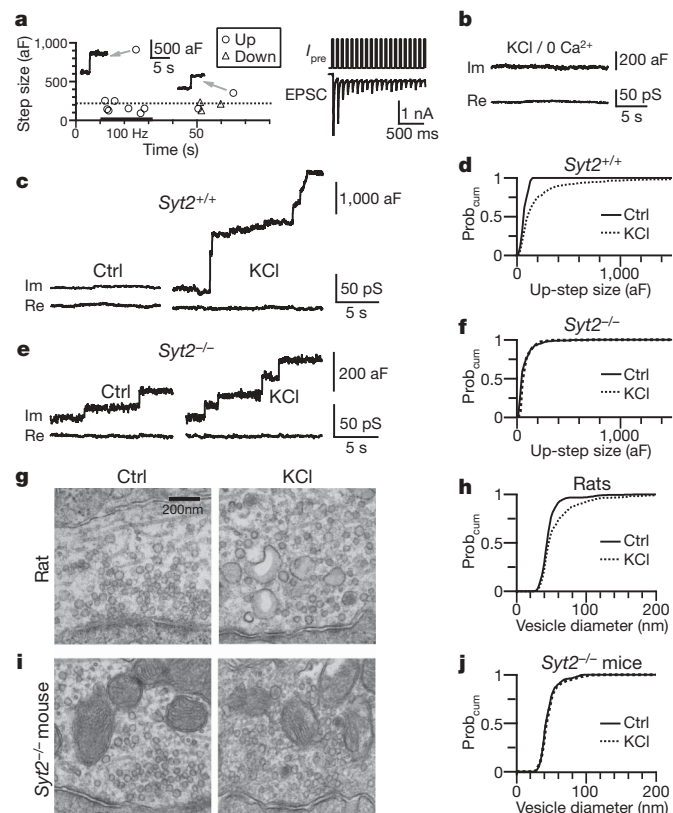
Endosome-like structures formed by bulk endocytosis may undergo exocytosis<sup>12</sup>. For two reasons, this mechanism did not underlie giant up-steps. First, the first giant up-step occurred  $\sim 80$  s earlier than the first giant down-step after KCl application ( $n = 17$  patches,  $P < 0.001$ , Fig. 1f, g), and was no later than the first regular-sized down-step (Supplementary Information II-3), suggesting that it was not caused by re-exocytosis of endocytic structures. Second, when calyces were whole-cell dialyzed with GDP $\beta$ S ( $0.3$  mM) to block endocytosis<sup>13</sup>, the up-step frequency ( $0.24 \pm 0.07$  Hz,  $n = 6$  patches) and the percentage ( $28\%$ ) of giant up-steps were similar to those without GDP $\beta$ S

<sup>1</sup>National Institute of Neurological Disorders and Stroke, 35 Convent Drive, Building 35, Room 2B-1012, Bethesda, Maryland 20892, USA. <sup>2</sup>Department of Pulmonary Medicine, The University of Texas M. D. Anderson Cancer Center, 2121 West Holcombe Boulevard, Box 1100, Houston, Texas 77030, USA.

\*These authors contributed equally to this work.



( $P > 0.08$ , for example, Fig. 1h), indicating that giant up-steps were not caused by re-exocytosis of endocytic structures. GDP $\beta$ S indeed blocked endocytosis, because only 1 down-step was observed within 300 s of KCl application in 6 cell-attached patches showing up-steps



(Fig. 1h); whereas without GDP $\beta$ S dialysis, down-steps were observed in all 17 patches showing up-steps ( $\chi^2$  test,  $P < 0.01$ ).

Similar to KCl application, action potential trains at 100 Hz for 10–20 s, as induced by current injection with another cell-attached electrode at the calyx (Fig. 2a,  $n = 5$ )<sup>14</sup>, increased the up-step frequency from  $0.011 \pm 0.011$  Hz (before stimulation) to  $0.465 \pm 0.078$  Hz during stimulation ( $n = 6$ ,  $P < 0.01$ ) and to  $0.040 \pm 0.018$  Hz within 50 s after stimulation. Action potential trains induced a large number of giant up-steps (9 out of 56 up-steps during, 5 out of 11 up-steps after, Fig. 2a).

Calcium binding with synaptotagmin II (Syt2) mediates evoked release at calyces<sup>14</sup>. Two pieces of evidence suggest that calcium/Syt2 mediates giant up-steps. First, after removing the extracellular calcium, giant up-steps were observed in 1 out of 21 patches during KCl application (Fig. 2b), which was significantly lower than control (17 out of 55 patches,  $\chi^2$  test,  $P < 0.01$ ). Thus, like regular up-steps<sup>6</sup>, calcium influx induced giant up-steps. Second, similar to results obtained in rats, in *Syt2*<sup>+/+</sup> mice, KCl increased the up-step frequency from  $0.003 \pm 0.001$  Hz to  $0.14 \pm 0.02$  Hz ( $n = 11$  patches,  $P < 0.001$ ), the amplitude from  $79 \pm 16$  aF ( $n = 26$ ) to  $216 \pm 9$  aF ( $n = 483$ ,  $P < 0.01$ ), and the giant up-step percentage to 24% (Fig. 2c, d). Compared to these results, *Syt2*<sup>-/-</sup> calyces showed a higher up-step frequency at rest

**Figure 2 | Calcium/Syt2 mediates compound fusion.** **a**, Left, up-steps were induced by current injection (1 nA, 3 ms, 100 Hz, 20 s, bar) via another cell-attached patch at the same calyx. Dotted line indicates 220 aF. Arrows indicate sampled up-steps. Right, calyx cell-attached current injection ( $I_{pre}$ ) induced calyx action potentials and thus EPSCs from a synapse without postsynaptic neuron removal<sup>14</sup>. **b**, Im and Re in KCl with no extracellular calcium. **c**, Up-steps in control and in KCl from a *Syt2*<sup>+/+</sup> mouse. **d**, The up-step amplitude  $Prob_{cum}$  in control (26 up-steps, 11 patches) and in KCl (483 up-steps, 11 patches) from 8 *Syt2*<sup>+/+</sup> mice. **e**, **f**, Similar to **c**, **d**, but from 9 *Syt2*<sup>-/-</sup> mice (control: 220 up-steps, 10 patches; KCl: 812 up-steps, 10 patches). **g**, EM images of rat calyx in control and in KCl. **h**, The vesicle diameter  $Prob_{cum}$  in control (2,896 vesicles) and in KCl (2,040 vesicles) from 6 rats. **i**, **j**, Similar to **g**, **h**, but from 4 *Syt2*<sup>-/-</sup> mice (control: 1,298 vesicles; KCl: 1,053 vesicles).



( $0.13 \pm 0.04$  Hz,  $n = 10$  patches,  $P < 0.01$ ) with a similar amplitude ( $80 \pm 5$  aF,  $n = 220$ ,  $P > 0.9$ , Fig. 2e, f). Only 5% of up-steps were  $>220$  aF in *Syt2*<sup>-/-</sup> mice at rest, similar to results obtained in rats. In *Syt2*<sup>-/-</sup> mice, KCl application slightly increased the up-step frequency to  $0.25 \pm 0.08$  Hz ( $n = 10$  patches,  $P < 0.05$ ) without increasing the mean up-step amplitude ( $87 \pm 3$  aF,  $n = 834$ ,  $P > 0.2$ ; Fig. 2e, f) or the giant up-step percentage (3.8%,  $\chi^2$  test,  $P = 0.26$ ). Thus, Syt2 mediated most KCl-induced up-steps, including giant ones.

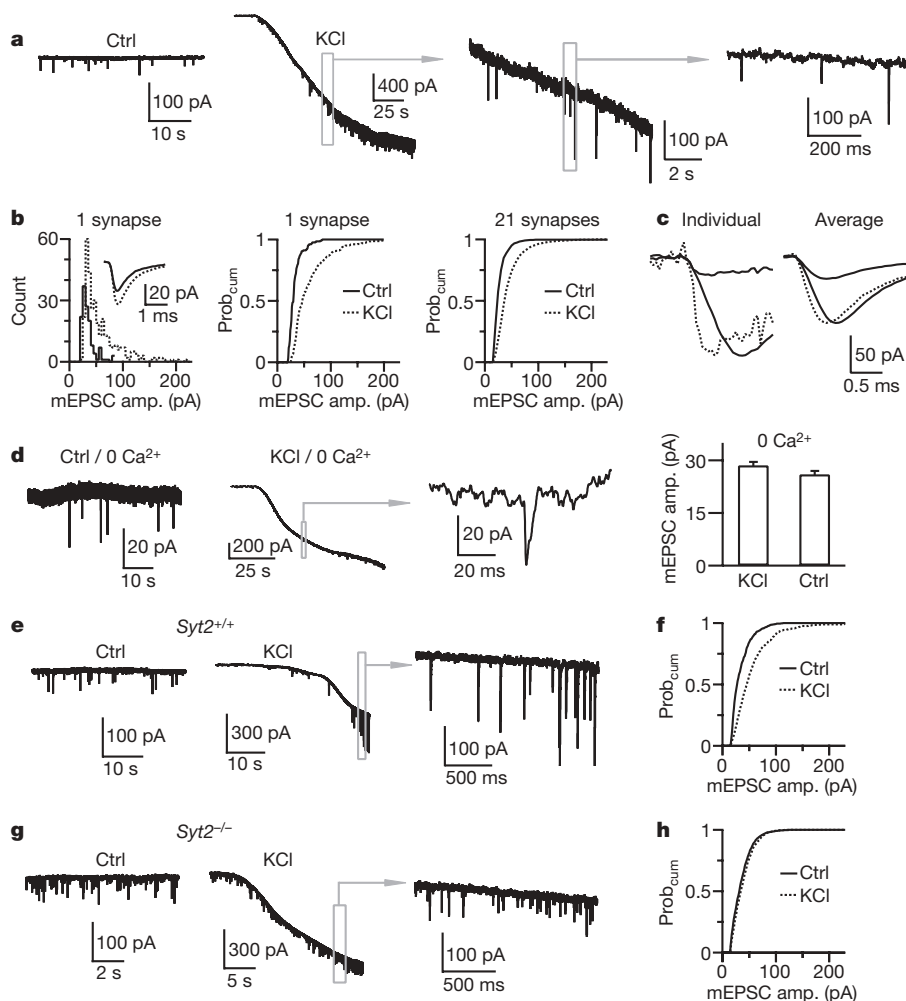
If calcium/Syt2-mediated giant up-steps reflect exocytosis of compound vesicles produced by compound fusion between vesicles<sup>2,3</sup>, KCl must induce Syt2-dependent giant vesicles and giant miniature EPSCs (mEPSCs). These predictions were tested as follows. First, at 1 min after KCl application, a time when giant down-steps reflecting bulk endocytosis rarely occurred (Fig. 1f, g), electron microscopy (EM) revealed that the vesicle diameter increased from  $47.2 \pm 0.4$  nm ( $n = 2,896$  vesicles) in control to  $57.8 \pm 0.9$  nm ( $n = 2,040$  vesicles from 6 rats,  $P < 0.001$ , Fig. 2g, h). About 13% of vesicles had a diameter from 80 to  $>300$  nm. These giant vesicles may have a membrane capacitance of  $\sim 220$  aF to  $>2,550$  aF (ref. 8), agreeing with the giant up-step amplitude range (220–2,878 aF, Fig. 1b). Similar results were observed at 1.5–2 min after KCl application in 4 rats and 4 *Syt2*<sup>+/+</sup> mice. Our observation of large organelles (see also ref. 10) made the occurrence of compound fusion physically possible. In *Syt2*<sup>-/-</sup> mice, however, 1.5 min of KCl application did not increase the vesicle diameter, supporting the suggestion that Syt2 mediates compound fusion (control:  $46.3 \pm 0.3$  nm,  $n = 1,298$  vesicles, 4 mice; KCl:  $47 \pm 0.5$  nm,  $n = 1,053$  vesicles, 4 mice,  $P > 0.3$ , Fig. 2i, j).

Second, KCl application for 2–3 min increased the mEPSC frequency from  $0.72 \pm 0.18$  Hz to  $6.37 \pm 0.6$  Hz ( $n = 21$  neurons,  $P < 0.01$ ), and the amplitude from  $29.6 \pm 1.0$  pA to  $44.9 \pm 2.3$  pA

( $n = 21$ ,  $P < 0.01$ , Fig. 3a, b). The fraction of mEPSCs larger than a value between 50 pA and 200 pA was higher during KCl application than in control ( $n = 21$  neurons,  $P < 0.01$ , Fig. 3b). The 10–90% rise time of mEPSCs  $>80$  pA ( $0.35 \pm 0.01$  ms) was slower than that for mEPSCs  $<80$  pA ( $0.29 \pm 0.01$  ms,  $n = 10$  neurons,  $P < 0.01$ , Fig. 3c), consistent with simulation showing that larger vesicles generate larger and slower mEPSCs (Supplementary Information III-1). The mEPSC amplitude increase was due to a larger amount of released glutamate (Supplementary Information III-2), consistent with exocytosis of giant vesicles. AMPA receptor saturation by higher glutamate concentrations might explain why KCl increased the mEPSC amplitude by only  $\sim 52\%$ , but increased the up-step size by  $\sim 103\%$  (Supplementary Information III-3).

Two pieces of evidence indicated that calcium/Syt2 was required for KCl-induced increases of the mEPSC frequency and amplitude. First, removing extracellular calcium abolished KCl-induced increase of the mEPSC frequency and amplitude ( $n = 5$  cells, Fig. 3d). Second, in *Syt2*<sup>+/+</sup> mice ( $n = 4$ ), KCl significantly increased the mEPSC frequency (control:  $0.7 \pm 0.3$  Hz; KCl:  $7.8 \pm 0.7$  Hz,  $n = 6$  synapses,  $P < 0.01$ , Fig. 3e) and amplitude (control:  $31.8 \pm 1.4$  pA; KCl:  $46.6 \pm 3.7$  pA,  $n = 6$ ,  $P < 0.01$ , Fig. 3f), whereas in *Syt2*<sup>-/-</sup> mice ( $n = 6$ ), KCl only slightly increased the mEPSC frequency (control:  $13.9 \pm 3.1$  Hz; KCl:  $18.8 \pm 2.9$  Hz,  $n = 9$ ,  $P < 0.01$ , Fig. 3g), but not the mEPSC amplitude (control:  $34.6 \pm 1.1$  pA; KCl:  $35.3 \pm 0.9$  pA,  $n = 9$ ,  $P > 0.3$ , Fig. 3h).

Giant up-steps after action potential trains (Fig. 2a) predicted giant mEPSCs. Indeed, action potential trains at 100 Hz for 4, 10 or 30 s increased the mEPSC amplitude, and induced post-tetanic potentiation (PTP) of the EPSC<sup>15,16</sup> (Fig. 4a, b). Longer trains induced larger increases (Fig. 4a, b). The PTP decay was fitted with a bi-exponential function, whereas the mEPSC amplitude potentiation decayed



**Figure 3 | KCl induces calcium/Syt2-dependent increase of the mEPSC size.** **a**, mEPSCs in control and in KCl (traces in boxes are enlarged). **b**, The distribution and the Prob<sub>cum</sub> for the mEPSC amplitude (amp.) in control and in KCl from 1 (inset shows the mean) and 21 synapses (control: 4,302 mEPSCs; KCl: 16,006 mEPSCs). **c**, Left, a small and a large mEPSC (solid), and the scaled small mEPSC (dotted). Right, the mean of mEPSCs  $<80$  pA and  $\geq 80$  pA (solid), and the scaled small mEPSC (dotted) from 1 synapse. **d**, Sampled mEPSCs and the mEPSC amplitude ( $\pm$  s.e.m., 5 neurons) in control and in KCl with no extracellular calcium. Right, 270 mEPSCs in control; 324 mEPSCs in KCl ( $P > 0.3$ ). **e**, mEPSCs in control and in KCl from a *Syt2*<sup>+/+</sup> mouse. **f**, The mEPSC amplitude Prob<sub>cum</sub> in control (1,788 mEPSCs, 6 synapses) and in KCl (3,360 mEPSCs, 6 synapses) from 4 *Syt2*<sup>+/+</sup> mice. **g**, **h**, Similar to **e**, **f**, but from 6 *Syt2*<sup>-/-</sup> mice (control: 3,780 mEPSCs, 9 synapses; KCl: 8,224 mEPSCs, 9 synapses).

approximately mono-exponentially (Fig. 4a). The amplitude and the time constant of the mEPSC amplitude potentiation matched the PTP slow component (Fig. 4b,  $P > 0.3$ ), suggesting a causal relation between them. The mEPSC size increase was small after 4 s stimulation ( $10 \pm 3\%$ ,  $n = 5$ , Fig. 4a), consistent with reports of negligible mEPSC size change<sup>15,16</sup>. Although these results were obtained in postnatal day (p)6–10 rats, similar results were obtained in p13–14 rats ( $n = 3$ ).

Bath application of the calcium buffer EGTA-AM (200  $\mu$ M,  $\sim 20$  min) reduced the EPSC amplitude<sup>15,16</sup> and abolished both the PTP<sup>15–17</sup> and the mEPSC size increase induced by a 30 s train ( $n = 4$ , Fig. 4c). Deletion of Syt2 in mice decreased the EPSC<sup>14</sup>, and abolished both PTP and the mEPSC amplitude potentiation (Fig. 4d, e). Syt2 deletion also reduced the mEPSC frequency increase, as measured in the first minute after stimulation ( $2.8 \pm 0.5$  times from 7 synapses in *Syt2*<sup>-/-</sup> mice, but  $11.5 \pm 4.2$  times from 7 synapses in *Syt2*<sup>+/-</sup> littermates,  $P < 0.01$ ), suggesting the involvement of Syt2 in the mEPSC frequency increase. Taken together, our results suggest that calcium/Syt2 mediates the mEPSC size increase, which generates the PTP slow component.

In summary, KCl application induced calcium/Syt2-dependent increases of up-steps (see also Supplementary Information IV), vesicle size, and the mEPSC amplitude and rise time, suggesting that calcium/Syt2 mediates compound fusion, which forms compound vesicles. Subsequent exocytosis of compound vesicles increased the quantal size, which contributed to the generation of PTP. These results suggest the need to include compound fusion between vesicles, along with vesicle fusion with the plasma membrane<sup>1</sup>, into the current model of exocytosis at synapses (Supplementary Fig. 1).

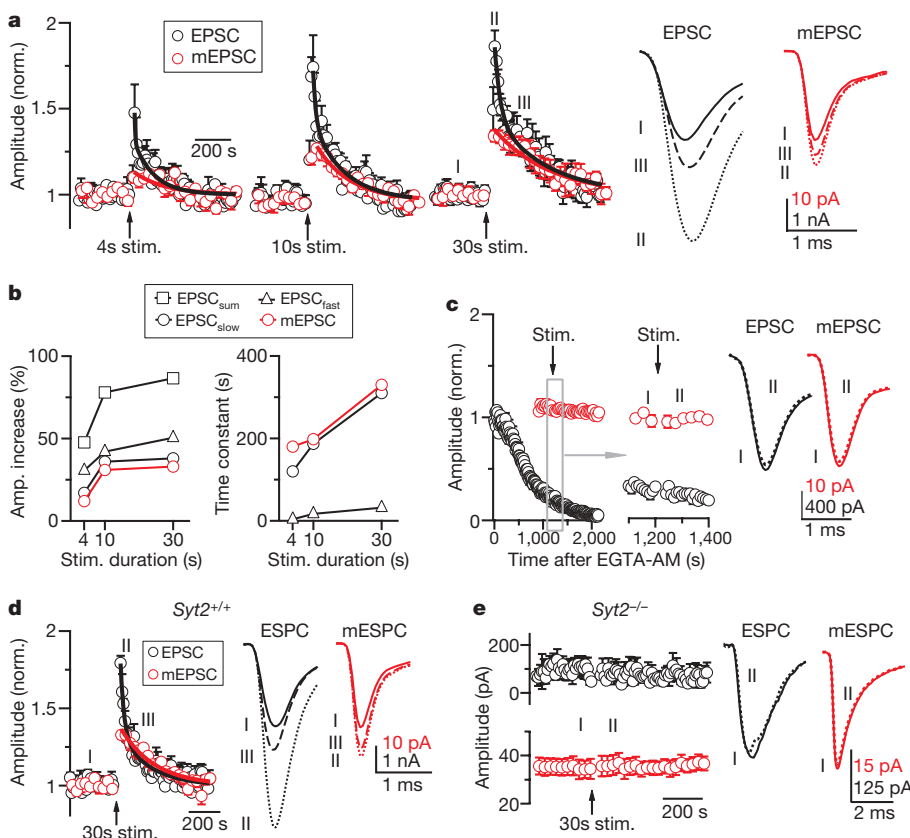
Giant miniature currents or potentials ('minis') observed at many synapses<sup>18–21</sup> are inconsistent with the quantal hypothesis, which assumes a single quantum for a mini<sup>22</sup>. Accordingly, multi-vesicular release was proposed<sup>18–21</sup>. Our results suggest that giant minis are caused by compound fusion and subsequent compound exocytosis. Consistent with this suggestion, giant organelles are present at synapses displaying giant minis<sup>18,19</sup>.

Multi-vesicular release was observed during stimulation at many synapses<sup>23–26</sup>. An alternative explanation to multi-vesicular release, compound fusion, was proposed<sup>4</sup>, and supported by a study revealing large organelles soon after  $\sim 20$  s stimulation at ribbon-type synapses, although bulk endocytosis as the cause of these large organelles was not fully excluded<sup>5</sup>. The present work suggests that compound fusion is an alternative explanation to multi-vesicular release at conventional synapses.

Large endosome-like structures were previously attributed to endocytosis<sup>9–12,27</sup>. They can now be reinterpreted at least partly as compound fusion (see also ref. 5). We found that giant down-steps reflecting bulk endocytosis occurred later than giant up-steps, and correlated with giant up-steps in frequency and size (Supplementary Information V). We suggest that bulk endocytosis retrieves exocytosed compound vesicles via a mechanism that maintains the identity or at least the area of compound vesicles (Supplementary Fig. 1).

PTP, which may control the neuronal circuit function<sup>17</sup>, was not considered to be caused by the quantal size increase, probably because at its peak, the mEPSC size increase was insufficient to account for PTP (Fig. 4a). The present work suggests that compound-fusion-induced quantal size increase contributes to the generation of PTP. Compound fusion is likely to be a common mechanism used to enhance synaptic strength, because (1) PTP is observed at many synapses<sup>15,17</sup>, (2) intense firing that may generate PTP occurs in physiological and pathological conditions<sup>17,28,29</sup>, and (3) large organelles<sup>9–12,27</sup> and giant mEPSCs<sup>18–21</sup> that may be caused by compound fusion are widely observed.

Although compound fusion was found in non-neuronal secretory cells decades ago<sup>2,3</sup>, its underlying mechanism is unclear. The present work suggests that calcium binding with synaptotagmin mediates compound fusion, probably via interaction between v- and t-SNARE fusion proteins on the vesicle membrane<sup>30</sup>. The affinity of Syt2 for calcium is at tens of micromolar<sup>14</sup>. The calcium micro-domain near calcium channels, in which the calcium concentration transiently increases to tens of micromolar during an action potential, might expand greatly during prolonged stimulation, allowing



Syt2 to mediate fusion between vesicles beyond the plasma membrane. Consistent with this possibility, EGTA blocked the mEPSC size increase (Fig. 4c), probably by limiting the expansion of the calcium microdomain during high frequency stimulation.

## METHODS SUMMARY

For most experiments, parasagittal brainstem slices (200  $\mu$ m thick) containing the medial nucleus of the trapezoid body were prepared from Wistar rats or mice (*Syt2*<sup>-/-</sup> or *Syt2*<sup>+/-</sup> littermates) using a vibratome<sup>6</sup>. For PTP experiments only, horizontal brainstem slices were prepared, where a 0.1 ms, 2–20 V voltage pulse was applied via a bipolar electrode at the midline of the trapezoid body to evoke action potentials at the axon and calyx. If not mentioned, animals were 6–10 days old. We removed the postsynaptic neuron and performed cell-attached recordings of the imaginary (Im, capacitance) and the real (Re, conductance) components of the complex admittance at the release face of the calyces of Held<sup>6</sup>. Except for Fig. 1a (right), Im and Re traces in all figures are low-pass filtered at 50 Hz. For traces in Fig. 1a (right), the low-pass filter in the lock-in amplifier was set at 0.1 ms. Means are presented as  $\pm$ s.e.m. Unless mentioned, the statistical test was the *t*-test. Detailed methods are described mostly in Methods, and partly in Supplementary Information VI.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 10 November 2008; accepted 3 February 2009.**

**Published online 11 March 2009; corrected 7 May 2009 (see full-text HTML version for details).**

- Sudhof, T. C. The synaptic vesicle cycle. *Annu. Rev. Neurosci.* **27**, 509–547 (2004).
- Alvarez, d. T. & Fernandez, J. M. Compound versus multigranular exocytosis in peritoneal mast cells. *J. Gen. Physiol.* **95**, 397–409 (1990).
- Scepek, S. & Lindau, M. Focal exocytosis by eosinophils—compound exocytosis and cumulative fusion. *EMBO J.* **12**, 1811–1817 (1993).
- Heidelberger, R., Heinemann, C., Neher, E. & Matthews, G. Calcium dependence of the rate of exocytosis in a synaptic terminal. *Nature* **371**, 513–515 (1994).
- Matthews, G. & Sterling, P. Evidence that vesicles undergo compound fusion on the synaptic ribbon. *J. Neurosci.* **28**, 5403–5411 (2008).
- He, L., Wu, X. S., Mohan, R. & Wu, L. G. Two modes of fusion pore opening revealed by cell-attached recordings at a synapse. *Nature* **444**, 102–105 (2006).
- Sätzler, K. *et al.* Three-dimensional reconstruction of a calyx of Held and its postsynaptic principal neuron in the medial nucleus of the trapezoid body. *J. Neurosci.* **22**, 10567–10579 (2002).
- Gentet, L. J., Stuart, G. J. & Clements, J. D. Direct measurement of specific membrane capacitance in neurons. *Biophys. J.* **79**, 314–320 (2000).
- Heuser, J. E. & Reese, T. S. Evidence for recycling of synaptic vesicle membrane during transmitter release at the frog neuromuscular junction. *J. Cell Biol.* **57**, 315–344 (1973).
- de Lange, R. P., de Roos, A. D. & Borst, J. G. Two modes of vesicle recycling in the rat calyx of Held. *J. Neurosci.* **23**, 10164–10173 (2003).
- Richards, D. A., Guatimosim, C. & Betz, W. J. Two endocytic recycling routes selectively fill two vesicle pools in frog motor nerve terminals. *Neuron* **27**, 551–559 (2000).
- Coggins, M. R., Grabner, C. P., Almers, W. & Zenisek, D. Stimulated exocytosis of endosomes in goldfish retinal bipolar neurons. *J. Physiol. (Lond.)* **584**, 853–865 (2007).
- Xu, J. *et al.* GTP-independent rapid and slow endocytosis at a central synapse. *Nature Neurosci.* **11**, 45–53 (2008).
- Sun, J. *et al.* A dual-Ca<sup>2+</sup>-sensor model for neurotransmitter release in a central synapse. *Nature* **450**, 676–682 (2007).
- Korogod, N., Lou, X. & Schneggenburger, R. Presynaptic Ca<sup>2+</sup> requirements and developmental regulation of posttetanic potentiation at the calyx of Held. *J. Neurosci.* **25**, 5127–5137 (2005).
- Habets, R. L. & Borst, J. G. Post-tetanic potentiation in the rat calyx of Held synapse. *J. Physiol. (Lond.)* **564**, 173–187 (2005).
- Zucker, R. S. & Regehr, W. G. Short-term synaptic plasticity. *Annu. Rev. Physiol.* **64**, 355–405 (2002).
- Heuser, J. E. A possible origin of the 'giant' spontaneous potentials that occur after prolonged transmitter release at frog neuromuscular junctions. *J. Physiol. (Lond.)* **239**, 106P–108P (1974).
- Henze, D. A., McMahon, D. B., Harris, K. M. & Barrionuevo, G. Giant miniature EPSCs at the hippocampal mossy fiber to CA3 pyramidal cell synapse are monoquantal. *J. Neurophysiol.* **87**, 15–29 (2002).
- Llano, I. *et al.* Presynaptic calcium stores underlie large-amplitude miniature IPSCs and spontaneous calcium transients. *Nature Neurosci.* **3**, 1256–1265 (2000).
- Wall, M. J. & Usowicz, M. M. Development of the quantal properties of evoked and spontaneous synaptic currents at a brain synapse. *Nature Neurosci.* **1**, 675–682 (1998).
- Del Castillo, J. & Katz, B. Quantal components of the end-plate potential. *J. Physiol. (Lond.)* **124**, 560–573 (1954).
- Tong, G. & Jahr, C. E. Multivesicular release from excitatory synapses of cultured hippocampal neurons. *Neuron* **12**, 51–59 (1994).
- Wadiche, J. I. & Jahr, C. E. Multivesicular release at climbing fiber-Purkinje cell synapses. *Neuron* **32**, 301–313 (2001).
- Sun, J. Y. & Wu, L. G. Fast kinetics of exocytosis revealed by simultaneous measurements of presynaptic capacitance and postsynaptic currents at a central synapse. *Neuron* **30**, 171–182 (2001).
- Singer, J. H., Lassova, L., Vardi, N. & Diamond, J. S. Coordinated multivesicular release at a mammalian ribbon synapse. *Nature Neurosci.* **7**, 826–833 (2004).
- Paillart, C., Li, J., Matthews, G. & Sterling, P. Endocytosis and vesicle recycling at a ribbon synapse. *J. Neurosci.* **23**, 4092–4099 (2003).
- Oertel, D. The role of timing in the brain stem auditory nuclei of vertebrates. *Annu. Rev. Physiol.* **61**, 497–519 (1999).
- Boraud, T., Bezard, E., Bioulac, B. & Gross, C. E. From single extracellular unit recording in experimental and human Parkinsonism to the development of a functional concept of the role played by the basal ganglia in motor control. *Prog. Neurobiol.* **66**, 265–283 (2002).
- Takamori, S. *et al.* Molecular anatomy of a trafficking organelle. *Cell* **127**, 831–846 (2006).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank J. Diamond and K. Paradiso for comments on the manuscript, and S. Cheng, R. Azzam and V. Crocker for help in EM. This work was supported by the National Institute of Neurological Disorders and Stroke Intramural Research Program (L.-G.W.) and the American Heart Association (R.A.).

**Author Contributions** L.H. performed cell-attached recordings; L.X. performed EM work, and the mEPSC and EPSC recordings; J.X. and B.D.M. helped with some experiments; L.B. helped with EM work and maintained *Syt2*<sup>-/-</sup> mice; E.M. and R.A. generated the *Syt2*<sup>-/-</sup> mouse line; and L.-G.W. supervised the project and wrote the paper.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to L.-G.W. ([wul@ninds.nih.gov](mailto:wul@ninds.nih.gov)).



## METHODS

**Bath and pipette solutions.** For preparing slices, we used a solution containing (in mM): 95 NaCl, 25 NaHCO<sub>3</sub>, 25 glucose, 50 sucrose, 2.5 KCl, 1.25 NaH<sub>2</sub>PO<sub>4</sub>, 0.1 CaCl<sub>2</sub>, 3 MgCl<sub>2</sub>, 0.4 ascorbic acid, 3 *myo*-inositol, 2 sodium pyruvate (95% O<sub>2</sub>/5% CO<sub>2</sub>). Slices were incubated for 30 min at 37 °C and then held at room temperature (22–24 °C) for experiments in a solution containing (in mM): 125 NaCl, 2.5 KCl, 1 MgCl<sub>2</sub>, 2 CaCl<sub>2</sub>, 25 dextrose, 1.25 NaH<sub>2</sub>PO<sub>4</sub>, 0.4 ascorbic acid, 3 *myo*-inositol, 2 sodium pyruvate, 25 NaHCO<sub>3</sub>, 0.001 TTX (preventing action potentials), pH 7.4 when bubbled with 95% O<sub>2</sub>, 5% CO<sub>2</sub>. When KCl concentration was increased (to 50–100 mM, if not mentioned), NaCl concentration was decreased to maintain the same osmolarity.

For cell-attached recordings and current injection (Figs 1, 2), the pipette (4–6 MΩ) contained (in mM): 130 NaCl, 2.5 KCl, 2 CaCl<sub>2</sub>, 1 MgCl<sub>2</sub>, 10 HEPES, 20 TEA, pH 7.2, adjusted with NaOH (osmolarity was 310–320 mOsm).

For whole-cell dialysis of GDPβS (Fig. 1h), the pipette contained (in mM): 125 K-gluconate, 20 KCl, 10 Na<sub>2</sub>-phosphocreatine, 4 MgATP, 0.3 GDPβS, 10 HEPES, 0.05 BAPTA, pH 7.2, adjusted with KOH. The calyx was under the current-clamp mode (EPC-10 amplifier, HEKA electronics) while high potassium was applied to induce exocytosis.

For recordings of the postsynaptic AMPA receptor-mediated EPSC and mEPSC (Figs 3, 4), the pipette (2–3 MΩ) contained (in mM): 125 K-gluconate, 20 KCl, 4 MgATP, 10 Na<sub>2</sub>-phosphocreatine, 0.3 GTP, 10 HEPES, and 0.5 EGTA, pH 7.2, adjusted with KOH. D-APV (50 μM), bicuculline (10 μM) and strychnine (10 μM) were added to the bath solution to isolate AMPA receptor-mediated mEPSCs.

**Cell-attached and whole-cell recordings and data analysis.** Cell-attached capacitance measurements at the release face of the calyx of Held were performed with an SR830 2-phase lock-in amplifier (Stanford Research Systems) coupled to an EPC-8 patch-clamp amplifier (HEKA electronics)<sup>6</sup>. A 20 kHz, 200 mV r.m.s. sine wave was superimposed on a command potential of 0 mV. The in-phase (real, Re) and 90°-out-of-phase (imaginary, Im) current outputs of the lock-in amplifier were low-pass filtered at 0.3–1 ms (24 dB) in most experiments. In experiments requiring higher time resolution, signals were filtered at 0.1 ms, which, together with our sampling rate of 2–3.3 kHz, gave a time resolution of ~0.6–0.9 ms (Fig. 1a, right). Methods of cell-attached capacitance recordings, phase adjustment, detection and measurement of capacitance up- and down-steps have been described previously<sup>6</sup>.

The fusion or the fission pore conductance ( $G_p$ ) was calculated based on

$$G_p = (\Delta \text{Re}^2 + \Delta \text{Im}^2) / \Delta \text{Re} \quad (1)$$

where  $\Delta$  means the difference from baseline<sup>31</sup>. Based on the  $G_p$  value, we estimated the fusion or the fission pore diameter ( $D_p$ ) based on<sup>32</sup>

$$D_p = [4G_p\rho\lambda/\pi]^{0.5} \quad (2)$$

where  $\rho$  is the saline resistivity (100 Ω cm), and  $\lambda$ , the pore length, is taken as the length of a gap junction channel (15 nm). The fusion  $G_p$  usually reached a plateau level in ~10 ms after fusion, at which we measured the initial  $G_p$  value (Fig. 1c). It was followed by a gradual increase for ~10–230 ms before rapid  $G_p$  expansion occurred (Fig. 1c).

We estimated that a cell-attached patch at the calyx release face might contain a single to a few active zones<sup>6</sup>. During application of ~50–100 mM KCl solution, up-steps were observed in 17 out of 55 patches (Fig. 1a, b). The percentage of failure (38/55 or ~2/3) was similar to that obtained during application of 25 mM

KCl in our previous study<sup>6</sup>. Patch at the residual postsynaptic membrane or the non-release face of the calyx might explain why capacitance up-steps were detected in ~1/3 of patches<sup>6</sup>.

Only patches showing up-steps during stimulation were used for data analysis. Up- and down-steps were collected within 500 s after KCl application. In most experiments, we used 50–100 mM KCl solution for stimulation. Up- and down-steps, including giant ones, were analysed from 17 patches showing up- and down-steps during application of ~50–100 mM KCl (Fig. 1b, e, red). Up- and down-steps before KCl application in these 17 patches were not sufficient for plotting their amplitude distribution and cumulative probability curve. To collect more data in control (before KCl application), we included data from additional 61 patches showing up-steps during application of ~15–50 mM KCl solution, which we collected in search of the optimal condition for the generation of giant up-steps. Thus, the control data before KCl application (Fig. 1b, e, black) were collected from 78 patches showing up-steps in response to 15–100 mM KCl solution.

Voltage-clamp recordings of EPSC were performed with an EPC 10 amplifier<sup>33</sup>. The series resistance (<10 MΩ) was compensated by 85–95% when the EPSC recording was made.

**The mEPSC analysis.** The mEPSC was analysed with a mini program (Mini Analysis Program, Synaptosoft Inc.). KCl increased the frequency and the amplitude of mEPSCs. These effects were observed during the first 2–3 min of KCl application. We limited data analysis to this period of time for two reasons. First, the frequency of mEPSCs increased to more than 100 Hz as the potassium concentration reached the steady-state. With such a high frequency, the chance for two independent mEPSCs to occur at the same time increased, and the postsynaptic AMPA receptor desensitization might occur, which complicated data analysis. Second, as the potassium concentration approached the steady-state, the membrane resistance of the postsynaptic neuron decreased from  $148 \pm 9$  to  $17.5 \pm 3$  MΩ ( $n = 5$ ), which was approaching the access resistance (~7–15 MΩ) and thus could decrease the quality of the voltage-clamp. The reason for this decrease is unclear. Both the decrease in the membrane resistance and the change in the reversal potential of potassium channels may contribute to the increase in the inward current during the first 2 min of KCl application (Fig. 3a). Thus, we limited data analysis to within 2–3 min after KCl application, during which the membrane resistance was greater than ~50 MΩ. Because of a higher noise during KCl application, the threshold of mEPSC detection was set at 20 pA, which also applied to the control condition.

For recordings of mEPSCs during KCl application, we did not compensate the access resistance (<10 MΩ) in most experiments, which reduced the noise. The increase of the mEPSC during KCl application was not due to inadequate voltage-clamp for two reasons. First, the mEPSC amplitude was not increased by KCl application when the extracellular calcium was removed (Fig. 3d). Second, when we compensated the access resistance by 90%, high potassium application still increased the mEPSC amplitude from  $31.4 \pm 0.9$  pA (546 mEPSCs, 4 neurons) in control to  $46.2 \pm 1.8$  pA (2,147 mEPSCs, 4 neurons, not shown), similar to those obtained without the access resistance compensation (Fig. 3b).

**Additional methods.** Syt2 knockout mice and EM are described in Supplementary Information VI.

1. Lindau, M. & Alvarez de Toledo, G. The fusion pore. *Biochim. Biophys. Acta* **164**, 167–173 (2003).
2. Spruce, A. E., Breckenridge, L. J., Lee, A. K. & Almers, W. Properties of the fusion pore that forms during exocytosis of a mast cell secretory vesicle. *Neuron* **4**, 643–654 (1990).
3. Xu, J. & Wu, L. G. The decrease in the presynaptic calcium current is a major cause of short-term depression at a calyx-type synapse. *Neuron* **46**, 633–645 (2005).

## LETTERS

# Fused has evolved divergent roles in vertebrate Hedgehog signalling and motile ciliogenesis

Christopher W. Wilson<sup>1\*</sup>, Catherine T. Nguyen<sup>2\*</sup>, Miao-Hsueh Chen<sup>1</sup>, Jehn-Hsiahn Yang<sup>1†</sup>, Rhodora Gacayan<sup>1</sup>, Jie Huang<sup>2</sup>, Jau-Nian Chen<sup>2</sup> & Pao-Tien Chuang<sup>1</sup>

Hedgehog (Hh) signalling is essential for several aspects of embryogenesis<sup>1,2</sup>. In *Drosophila*, Hh transduction is mediated by a cytoplasmic signalling complex<sup>3–5</sup> that includes the putative serine-threonine kinase Fused (Fu) and the kinesin Costal 2 (Cos2, also known as Cos), yet Fu does not have a conserved role in Hh signalling in mammals<sup>6,7</sup>. Mouse *Fu* (also known as *Stk36*) mutants are viable and seem to respond normally to Hh signalling. Here we show that mouse *Fu* is essential for construction of the central pair apparatus of motile, 9+2 cilia and offers a new model of human primary ciliary dyskinesia. We found that mouse *Fu* physically interacts with Kif27, a mammalian Cos2 orthologue<sup>8</sup>, and linked *Fu* to known structural components of the central pair apparatus, providing evidence for the first regulatory component involved in central pair construction. We also demonstrated that zebrafish *Fu* is required both for Hh signalling and cilia biogenesis in Kupffer's vesicle. Mouse *Fu* rescued both Hh-dependent and -independent defects in zebrafish. Our results delineate a new pathway for central pair apparatus assembly, identify common regulators of Hh signalling and motile ciliogenesis, and provide insights into the evolution of the Hh cascade.

To further investigate the role of *Fu* in mammalian Hh signalling, we addressed whether Hh-dependent Smo localization to the primary cilium is affected in the absence of *Fu*. Primary cilia, which have a '9+0' arrangement of nine outer doublet microtubules, are required for Hh responses and contain several Hh pathway components<sup>2,9</sup>. We found that *Fu*<sup>-/-</sup> mouse embryonic fibroblasts formed primary cilia normally, trafficked Smo to the primary cilium in response to Hh ligand, and exhibited a typical Gli transcriptional response (Supplementary Figs 1, 2 and data not shown). This suggests that the single mammalian *Fu* orthologue is dispensable for Hh signalling. To explore the function of *Fu* in mice, we examined its expression in postnatal tissues by *in situ* hybridization. The *Fu* transcript was strongly expressed in the respiratory epithelium, the ependymal lining of the ventricles in the brain, and in the oviduct and testis (Fig. 1a–c and data not shown). These expression patterns are reminiscent of genes involved in the biogenesis of motile cilia, which function in these tissues to propel mucus, fluid and cells. In contrast to the primary cilium, the classical '9+2' motile cilium consists of nine outer doublet microtubules and two singlet central pair microtubules<sup>10</sup>. The central pair apparatus has a crucial involvement in regulating ciliary motility, but its formation is poorly understood because the centriole-derived basal body, from which the cilia axoneme extends, does not provide a template for central pair outgrowth. Disruption of human motile cilia function leads to primary ciliary dyskinesia, which is associated with recurrent respiratory infection, hydrocephalus and infertility<sup>11–13</sup>. To determine whether motile cilia function is compromised in *Fu*<sup>-/-</sup>

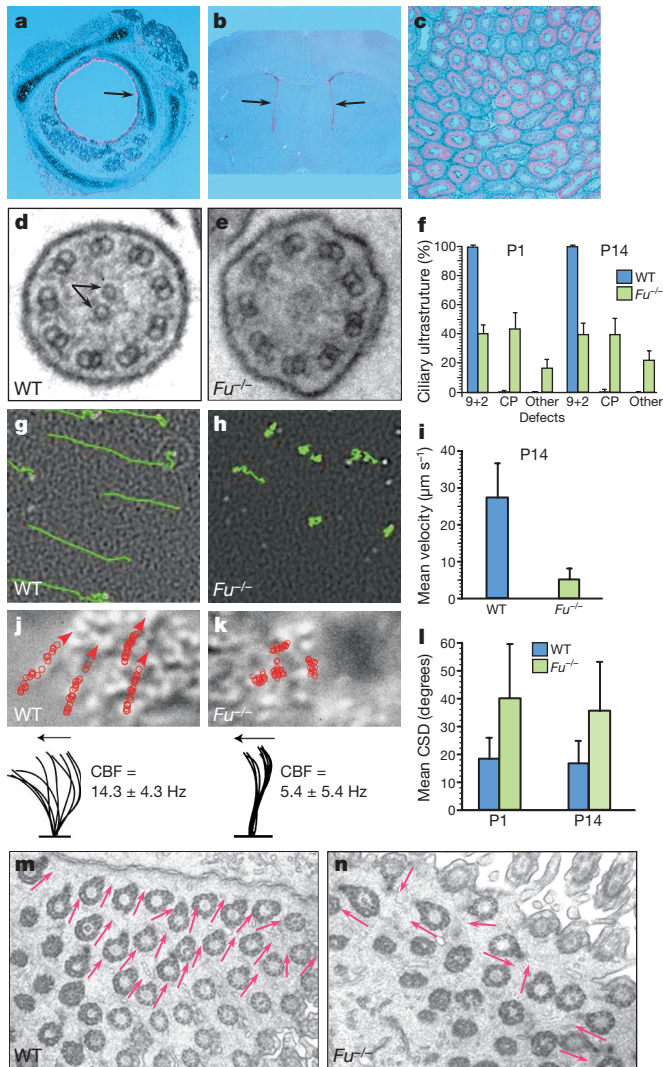
mice, we studied cilia axonemal ultrastructure by transmission electron microscopy. In wild-type animals, over 99% of tracheal and ependymal motile cilia showed a typical 9+2 configuration (Fig. 1d, f and Supplementary Figs 3 and 4). In contrast, approximately 60% of *Fu*<sup>-/-</sup> cilia have abnormal ciliary ultrastructure; two-thirds of which lack the central pair apparatus (Fig. 1e, f and Supplementary Figs 3 and 4). Central pair defects in *Fu* mutants are apparent at the basal plate region characterized by an electron-dense thick band, where the central pair originates in wild-type cilia<sup>14</sup> (Supplementary Fig. 4). Our findings indicate that mammalian *Fu* is dispensable for Hh signalling and specifically participates in the generation of the central pair apparatus in motile cilia axonemes.

Mice lacking functional *Fu* are born with no obvious phenotype, but they fail to thrive in comparison with wild-type or heterozygous littermates and die before postnatal day (P) 21 (refs 6, 7). To determine the functional consequences of central pair apparatus loss in *Fu*<sup>-/-</sup> animals, we examined fluid flow in tracheal explants. Analysis of fluorescent bead movement showed strong distal–proximal directional flow in wild-type explants, whereas beads overlaid on *Fu*<sup>-/-</sup> tracheae showed severely impaired velocity and little to no directional movement (Fig. 1g–i and Supplementary Movies 1 and 2). We next determined whether elimination of the central pair apparatus in *Fu*<sup>-/-</sup> animals disrupted cilia motion. In wild-type tracheae, cilia beat in a linear path with a quick forward power stroke and a slower recovery stroke<sup>15</sup> (Fig. 1j and Supplementary Movies 3 and 5). Most *Fu*<sup>-/-</sup> cilia moved stiffly and had a markedly reduced stroke amplitude; a subset were either immotile or beat in a slow, circular motion (Fig. 1k and Supplementary Movies 4 and 6). In contrast to wild-type motile cilia, which beat coordinately to produce a metachronic wave<sup>16</sup>, cilia in *Fu*<sup>-/-</sup> animals that beat seemed disoriented with respect to their neighbours (Fig. 1k and Supplementary Movies 4 and 6). This prompted us to investigate whether cilia orientation, specified by a basal body accessory structure known as the basal foot, was perturbed<sup>17,18</sup>. In wild-type tracheae, basal feet were properly aligned with each other (Fig. 1m). In *Fu*<sup>-/-</sup> mutants, basal feet were disoriented and frequently pointed at right angles or antiparallel to one another (Fig. 1n), and the circular standard deviation of cilia orientation within a given cell was significantly higher (Fig. 1l). Loss of the central pair apparatus in *Fu*<sup>-/-</sup> mice thus eliminated directional fluid flow, resulting from uncoordinated ciliary beating and global disorganization of cilia polarity.

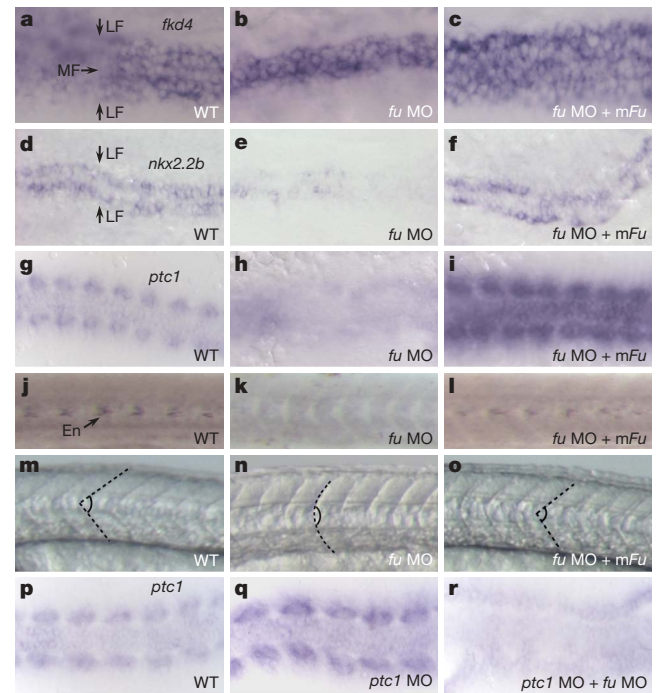
We proposed that *Fu* in different metazoan species might participate in both Hh signalling and ciliogenesis. We examined the role of *Fu* in zebrafish because *fu* (also known as *stk36*) morphants exhibit mild Hh-dependent somite phenotypes<sup>19</sup>. By delivering a higher concentration of *fu* morpholino, we observed stronger Hh phenotypes, including

<sup>1</sup>Cardiovascular Research Institute, University of California, San Francisco, California 94158, USA. <sup>2</sup>Department of Molecular, Cell and Developmental Biology, University of California, Los Angeles, California 90095, USA. <sup>†</sup>Present address: Department of Obstetrics and Gynecology, College of Medicine and the Hospital, National Taiwan University, Taipei, Taiwan. \*These authors contributed equally to this work.





cyclopia and loss of lateral floor plate (Fig. 2a, b, d, e, Supplementary Fig. 5 and data not shown), similar to *smo* mutants<sup>20</sup>. Knockdown of zebrafish *fu* activity greatly reduced *patched1* (*ptc1*) expression in somites, suggesting disruption of Hh responses (Fig. 2g, h). The



Hh-dependent muscle pioneer population, marked by the expression of *engrailed 1a* and *1b* (*eng1a* and *eng1b*), was lost (Fig. 2j, k), and *fu* morphants developed U-shaped instead of chevron-shaped somites (Fig. 2m, n). In *ptc1* morphants, Hh target genes are upregulated cell autonomously (Fig. 2p, q and Supplementary Fig. 10)<sup>19</sup>. Upregulation of Hh target genes is abolished in *ptc1;fu* double morphants (Fig. 2r), indicating that *fu* functions cell autonomously in Hh-responsive cells to control Hh signalling. Taken together, these results provide convincing evidence for an integral role of *Fu* in the zebrafish Hh pathway. We then addressed whether mouse *Fu* compensated for loss of zebrafish *Fu*. Surprisingly, co-injection of mouse *Fu* messenger RNA and zebrafish *fu* morpholino rescued all Hh phenotypes, including restoration of *ptc1* expression, lateral floor plate formation, muscle pioneer differentiation and somite shape (Fig. 2c, f, i, l, o, Supplementary Fig. 5 and data not shown). In contrast, co-injection of *Drosophila fu* mRNA and the zebrafish *fu* morpholino failed to rescue Hh phenotypes (data not shown). Thus, mouse *Fu* retains the necessary information to participate in the fish Hh pathway, indicating that a common mechanism underlies critical aspects of Hh signalling and motile ciliogenesis.

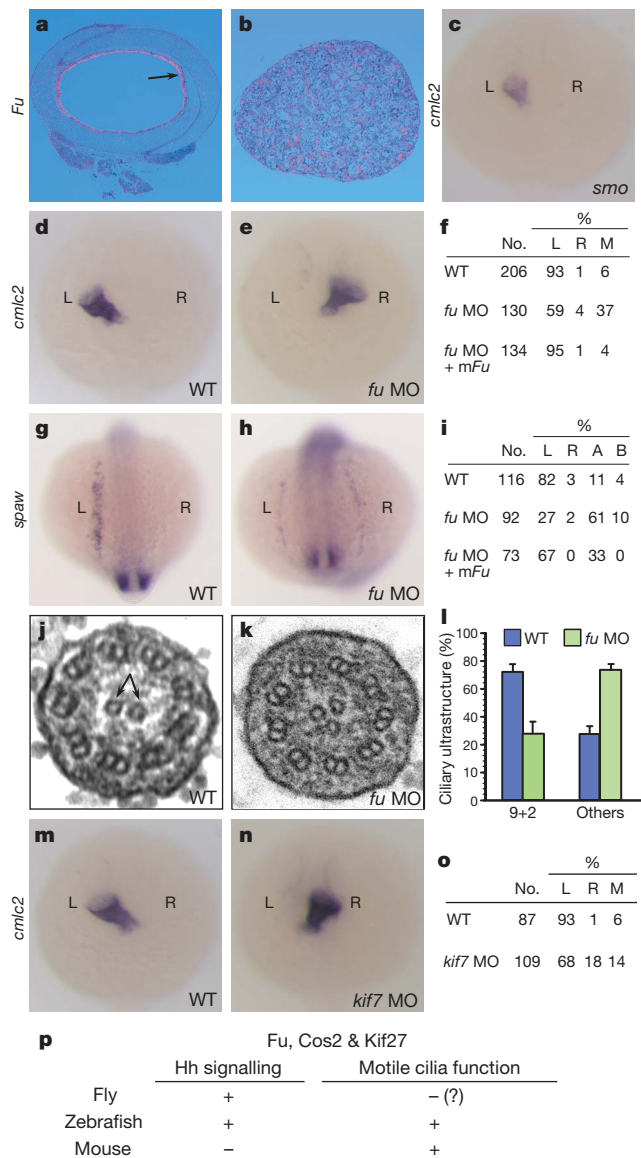


Fu may have an ancient, conserved role in regulating microtubule or motile cilia function because the genomes of many organisms, including plants and flagellated unicellular eukaryotes, contain genes encoding a highly conserved Fu kinase domain<sup>21</sup> (Supplementary Fig. 8). To test this idea, we examined *Fu* expression in different species by *in situ* hybridization and found strong expression in the chick tracheal epithelium and the oviduct and testis of *Xenopus tropicalis* (Fig. 3a, b and data not shown), in a pattern similar to mouse *Fu*. We then focused on zebrafish, which use 9+2 motile cilia on the surface of Kupffer's vesicle to generate an anticlockwise flow essential for establishment of left–right asymmetry<sup>22</sup>. If zebrafish *Fu* also participates in 9+2 cilia biogenesis, we reasoned that left–right asymmetry would be disrupted. We examined the positioning of the heart and visceral organs by cardiac myosin light chain 2 (*cmlc2*, also known as *myl7*) and fork head domain protein 2 (*fkd2*, also known as *foxa3*) expression, respectively. In contrast to zebrafish *smo* mutants, in which disrupted Hh signalling does not perturb left–right asymmetry<sup>20</sup> (Fig. 3c), 41% of *fu* morphants had reversed or midline hearts (Fig. 3d–f), whereas 30% of injected embryos had abnormal positioning of the gut, liver and pancreas (Supplementary Fig. 11). To investigate whether *Fu* is required for the early establishment of asymmetric gene expression in the left lateral plate mesoderm, we studied the expression pattern of *southpaw* (*spaw*) and paired-like homeodomain transcription factor 2 (*pitx2*) in *fu* morphants. In 73% of *fu* morphants, *spaw* was found to be on the right side, bilateral, or absent in the lateral plate mesoderm (Fig. 3g–i). Similarly, 71% of *fu* morphants had markedly reduced or absent *pitx2* staining in the lateral plate mesoderm (data not shown). Co-injection of mouse *Fu*, but not *Drosophila fu*, with *fu* morpholino was sufficient to restore left–right asymmetry (Fig. 3f, i and data not shown).

To confirm a direct role for *Fu* in regulating Kupffer's vesicle function, we injected fluorescein-labelled *fu* morpholino into dorsal forerunner cells<sup>23</sup>, which migrate at the leading edge of the embryonic shield to produce Kupffer's vesicle. Forty-four per cent of embryos with a strong fluorescent signal in the dorsal forerunner cells developed cardiac laterality but not somite defects (data not shown), indicating that the knockdown of *fu* in Kupffer's vesicle accounts for the left–right asymmetry defects. Kupffer's vesicle cilia in *fu* morphants had disorganized axonemal structures, including loss and acquisition of extra central pair microtubules (Fig. 3j–l and data not shown), indicating a conserved role of vertebrate *Fu* in central pair construction. Loss of *fu* affected cilia motility as shown by injecting rhodamine-conjugated dextran beads into Kupffer's vesicle of *fu* morphants at the 8-somite stage (Supplementary Movies 7 and 8). Defects in establishing an anticlockwise flow in *fu* morphants were rescued by mouse *Fu* (Supplementary Movie 9). Taken together, the data strongly support a conserved, Hh-independent role of *Fu* in vertebrate 9+2 cilia biogenesis (Fig. 3p).

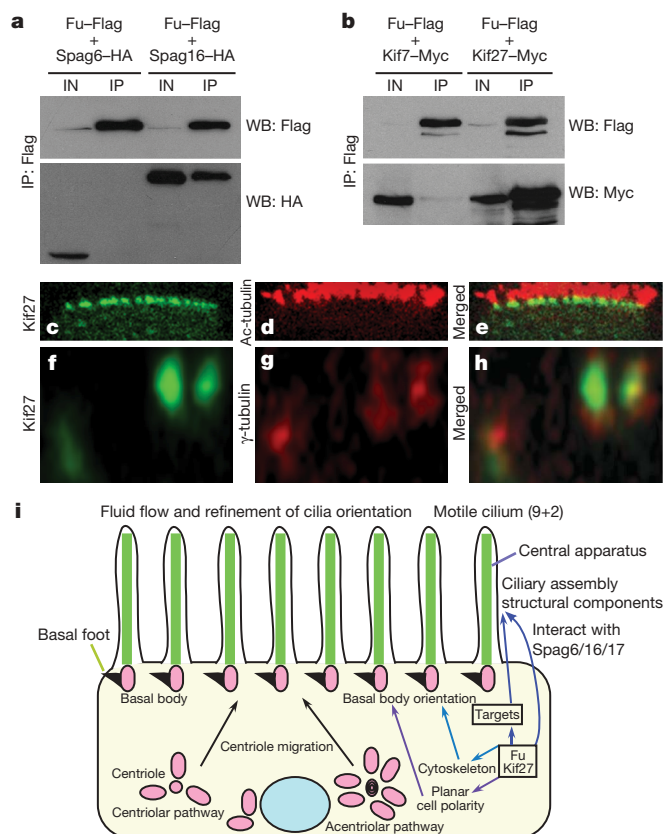
The process of central pair construction is poorly characterized and *Fu* is the first regulatory component known to control its assembly. To determine how *Fu* might control this process, we tested the ability of *Fu* to interact with Spag6 (also known as Pf16) and Spag16 (Pf20), evolutionarily conserved components of the central pair apparatus<sup>24,25</sup>. When expressed in HEK 293T cells, *Fu*–Flag efficiently co-immunoprecipitated Spag16–haemagglutinin (HA), but not Spag6–HA (Fig. 4a). Notably, Spag16 localizes to the sperm central pair apparatus<sup>26</sup>, and its *Chlamydomonas* orthologue Pf20 decorates the C2 microtubule along the intermicrotubule bridges between central pair microtubules<sup>27</sup>. This suggests a direct role for *Fu* in the assembly or maintenance of the central pair apparatus.

In fly, *Fu* binds to the kinesin Cos2 to transduce the Hh signal downstream of Smo. We examined whether mouse *Fu* bound to the mouse Cos2 orthologues Kif7 and Kif27. When expressed in HEK 293T cells and mouse tracheal epithelial cells (MTECs), *Fu*–Flag bound strongly to Kif27–Myc, but not to Kif7–Myc (Fig. 4b and data not shown), implicating Kif27 in the generation or regulation of 9+2 cilia. We expressed Kif27–green fluorescent protein (GFP) in MTECs by lentiviral infection and assessed its localization throughout MTEC



**Figure 3 | Zebrafish *fu* has an Hh-independent role in left–right asymmetry and generation of 9+2 cilia.** **a, b**, Section *in situ* hybridization to mouse *Fu* (pink signal) in chick trachea (**a**) and *X. tropicalis* testis (**b**). Arrow indicates sites of *Fu* expression. **c**, Whole mount *in situ* hybridization to *cmlc2* (purple signal) in *smo*<sup>hi1640Tg</sup> fish embryos at 24 h.p.f. View is dorsal. L, left; R, right. **d, e**, Whole-mount *in situ* hybridization to *cmlc2* in wild-type (**d**) and *fu* morphants (MO; **e**) at 24 h.p.f. View is dorsal. **f**, Summary of cardiac laterality defects in wild type ( $n = 206$ ), *fu* morphants ( $n = 130$ ), and *fu* morphants rescued with mouse *Fu* ( $n = 134$ ). M, medial. **g, h**, Whole mount *in situ* hybridization to *spaw* at the 15-somite stage. View is dorsal. **i**, Summary of *spaw* expression in the lateral plate mesoderm in wild type ( $n = 116$ ), *fu* morphants ( $n = 92$ ), and *fu* morphants rescued with mouse *Fu* ( $n = 73$ ). A, absent; B, bilateral. **j, k**, Electron micrograph of Kupffer's vesicle cilia from wild type (**j**) and a *fu* morphant (**k**). **l**, Quantification of ultrastructural defects in Kupffer's-vesicle cilia from wild type and *fu* morphants. Error bars indicate s.d. **m, n**, Whole mount *in situ* hybridization to *cmlc2* in wild-type and *kif7* morphants at 24 h.p.f. View is dorsal. **o**, Summary of cardiac laterality defects in wild-type ( $n = 87$ ) and *kif7* morphants ( $n = 109$ ). **p**, Summary of essential *Fu*, Cos2 and Kif27 functions in metazoan model organisms. Original magnification,  $\times 40$  (**a, b**),  $\times 80$  (**c–e, m, n**),  $\times 105.6$  (**g, h**) and  $\times 100,000$  (**j, k**).

differentiation induced by the creation of an air–liquid interface. During this process, hundreds of centrioles migrate to the apical surface of the cell, dock with the membrane to form basal bodies, and act as templates for the outgrowth of the outer microtubule doublets of the ciliary axoneme<sup>28</sup>. At air–liquid interface days 0 and



**Figure 4 | Mouse Fu interacts with the central pair protein Spag16 and the Cos2 orthologue Kif27.** **a**, Western blot of immunoprecipitated mouse Fu-Flag to detect its physical interaction with mouse Spag6-HA or Spag16-HA from HEK 293T lysates. IN, input; IP, immunoprecipitation; WB, western blot. **b**, Western blot of immunoprecipitated mouse Fu-Flag to determine its physical association with mouse Kif7-Myc or Kif27-Myc from HEK 293T lysates. **c–h**, Confocal images of fully differentiated MTECs to visualize localization of Kif27-GFP to the basal body (marked by anti- $\gamma$ -tubulin) of motile cilia labelled with acetylated tubulin (Ac-tubulin). **i**, Model of Fu, Kif27 and Spag16 function in motile cilia construction. Original magnification,  $\times 1,500$  (**c–h**).

5, Kif27-GFP punctae were associated with centrioles as determined by  $\gamma$ -tubulin staining (Supplementary Fig. 6). Kif27-GFP associated with the base of the cilium after axoneme outgrowth (Fig. 4c–h). Fu-mCherry was broadly distributed in the cytoplasm of MTECs throughout differentiation, overlapping with Kif27 (Supplementary Fig. 6). Fu and Kif27 expression are upregulated during MTEC differentiation, consistent with their essential roles in motile ciliogenesis (Supplementary Fig. 7). Efforts to demonstrate Fu kinase activity *in vitro* have not been successful, suggesting the requirement of a special microenvironment for its activity. We speculate that Fu has several substrates, some of which could reside in the cytoplasm and control central pair assembly indirectly (Fig. 4i). Our data favour a model in which Kif27 and/or Spag16 directs the localization or activity of Fu for central pair construction (Fig. 4i).

Despite the non-essential role of Fu in mammalian Hh signalling, the protein retains an interaction with the Cos2 orthologue Kif27. Analysis of Cos2, Kif7 and Kif27 sequences indicates that the *Kif7* and *Kif27* genes may have arisen by a duplication event (Supplementary Fig. 9). The four fish species examined do not contain an obvious Kif27 orthologue, suggesting either that Kif27 was lost after gene duplication, or that the duplication event occurred after divergence of the fish and amphibian lineages. Supporting the latter, morpholino knockdown of *kif7* in zebrafish (*z*) resulted in both Hh-specific phenotypes and disruption of left–right asymmetry (Fig. 3m–o and data not shown), indicating a dual role for Kif7 in Hh signalling<sup>29</sup> and

motile ciliogenesis, similar to zFu which co-immunoprecipitates with zKif7 (Supplementary Fig. 12). There are conflicting reports on the roles of Kif7 and Kif27 in vertebrate Hh signalling<sup>29,30</sup>; on the basis of data here, we predict that Kif27 does not have a vital role in mammalian Hh signal transduction, and mice lacking functional Kif27 may have phenotypes similar to Fu. We speculate that Fu has evolved or retained its function in central pair assembly in vertebrates, and that duplication of ancestral Cos2 in the vertebrate lineage led to the partition of functions for Kif7 and Kif27, while Kif27 retained its partnership with Fu (Fig. 3p). Although the requirement of Fu-like activity in mammalian Hh signalling is unproven, if it exists it is probably compensated for by an unrelated kinase (Supplementary Fig. 8). Alternatively, the involvement of the primary cilium as a scaffold for Hh pathway components in mammals could circumvent the need for a Fu–kinesin complex. Consistent with the notion of evolutionary changes in Hh pathway design in different species, Su(fu), a component of the cytoplasmic signalling complex, is dispensable for fly viability, plays a minor role in zebrafish Hh signalling and becomes an important negative regulator in mice<sup>2</sup>. Further analysis of Fu and Kif27 function in ciliogenesis and Hh signalling in diverse species will provide further insight into the evolution of this critical signalling pathway.

## METHODS SUMMARY

**Transmission electron microscopy.** Mouse tissue was fixed in 3% glutaraldehyde, 1% paraformaldehyde, 0.1 M sodium cacodylate, pH 7.4, at 4 °C overnight. Fish embryos were fixed in 2% paraformaldehyde, 2% glutaraldehyde (electron-microscopy grade) at room temperature for 2 h. Standard processing, embedding and sectioning procedures were followed. Samples were examined on a JEOL 100CX or JEM-1230 transmission electron microscope.

**Basal foot polarity.** The orientation and circular standard deviation of basal feet in electron microscopy micrographs was calculated as described<sup>18</sup>. Circular statistics were calculated using Oriana 2.0 (Kovach Computing Services).

**Tracheal flow assays.** Tracheae from P14 wild-type and *Fu*<sup>−/−</sup> mice were excised, cleaned of muscle and vasculature, opened longitudinally, and placed in a drop of PBS on a glass slide. Five microlitres of a 0.01% solution of Fluospheres (Invitrogen) were added on top of a single trachea to visualize the direction of ciliary flow. Images were acquired using a SPOT 2.3 camera connected to a Nikon E1000 epifluorescence microscope. Images were captured at a rate of 26 frames per second (f.p.s.) over a 50  $\mu$ m  $\times$  50  $\mu$ m area and were saved as .tiff stacks. Movies were examined in NIH Image J using the enhancing feature of the SpotTracker plugin (D. Sage and S. Gasser) to optimize sphere intensity, and the MtrackJ plugin (E. Meijering, Biomedical Imaging Group, University Medical Center, Rotterdam) to trace the direction and path length of the sphere. Average velocity was taken to be the straight-line distance a particle travelled from its originating point divided by time, and was calculated in Microsoft Excel.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 10 September 2008; accepted 10 February 2009.

Published online 22 March 2009.

- McMahon, A. P., Ingham, P. W. & Tabin, C. J. Developmental roles and clinical significance of hedgehog signaling. *Curr. Top. Dev. Biol.* **53**, 1–114 (2003).
- Huangfu, D. & Anderson, K. V. Signaling from Smo to Ci/Gli: conservation and divergence of Hedgehog pathways from *Drosophila* to vertebrates. *Development* **133**, 3–14 (2006).
- Sisson, J. C., Ho, K. S., Suyama, K. & Scott, M. P. Costal2, a novel kinesin-related protein in the Hedgehog signaling pathway. *Cell* **90**, 235–245 (1997).
- Robbins, D. J. et al. Hedgehog elicits signal transduction by means of a large complex containing the kinesin-related protein costal2. *Cell* **90**, 225–234 (1997).
- Lum, L. et al. Hedgehog signal transduction via Smoothened association with a cytoplasmic complex scaffolded by the atypical kinesin, Costal-2. *Mol. Cell* **12**, 1261–1274 (2003).
- Chen, M. H., Gao, N., Kawakami, T. & Chuang, P. T. Mice deficient in the fused homolog do not exhibit phenotypes indicative of perturbed hedgehog signaling during embryonic development. *Mol. Cell. Biol.* **25**, 7042–7053 (2005).
- Merchant, M. et al. Loss of the serine/threonine kinase fused results in postnatal growth defects and lethality due to progressive hydrocephalus. *Mol. Cell. Biol.* **25**, 7054–7068 (2005).
- Katoh, Y. & Katoh, M. KIF27 is one of orthologs for *Drosophila* Costal-2. *Int. J. Oncol.* **25**, 1875–1880 (2004).

9. Davenport, J. R. & Yoder, B. K. An incredible decade for the primary cilium: a look at a once-forgotten organelle. *Am. J. Physiol. Renal Physiol.* **289**, F1159–F1169 (2005).
10. Davis, E. E., Brueckner, M. & Katsanis, N. The emerging complexity of the vertebrate cilium: new functional roles for an ancient organelle. *Dev. Cell* **11**, 9–19 (2006).
11. Afzelius, B. A. Cilia-related diseases. *J. Pathol.* **204**, 470–477 (2004).
12. Marshall, W. F. & Kintner, C. Cilia orientation and the fluid mechanics of development. *Curr. Opin. Cell Biol.* **20**, 48–52 (2008).
13. Zariwala, M. A., Knowles, M. R. & Omran, H. Genetic defects in ciliary structure and function. *Annu. Rev. Physiol.* **69**, 423–450 (2007).
14. McKean, P. G., Baines, A., Vaughan, S. & Gull, K.  $\gamma$ -Tubulin functions in the nucleation of a discrete subset of microtubules in the eukaryotic flagellum. *Curr. Biol.* **13**, 598–602 (2003).
15. Chilvers, M. A., Rutman, A. & O'Callaghan, C. Ciliary beat pattern is associated with specific ultrastructural defects in primary ciliary dyskinesia. *J. Allergy Clin. Immunol.* **112**, 518–524 (2003).
16. Yang, X., Dillon, R. H. & Fauci, L. J. An integrative computational model of multiciliary beating. *Bull. Math. Biol.* **70**, 1192–1215 (2008).
17. Frisch, D. & Farbman, A. I. Development of order during ciliogenesis. *Anat. Rec.* **162**, 221–232 (1968).
18. Mitchell, B., Jacobs, R., Li, J., Chien, S. & Kintner, C. A positive feedback mechanism governs the polarity and motion of motile cilia. *Nature* **447**, 97–101 (2007).
19. Wolff, C., Roy, S. & Ingham, P. W. Multiple muscle cell identities induced by distinct levels and timing of hedgehog activity in the zebrafish embryo. *Curr. Biol.* **13**, 1169–1181 (2003).
20. Chen, W., Burgess, S. & Hopkins, N. Analysis of the zebrafish smoothened mutant reveals conserved and divergent functions of hedgehog activity. *Development* **128**, 2385–2396 (2001).
21. Oh, S. A. *et al.* A divergent cellular role for the FUSED kinase family in the plant-specific cytokinetic phragmoplast. *Curr. Biol.* **15**, 2107–2111 (2005).
22. Kramer-Zucker, A. G. *et al.* Cilia-driven fluid flow in the zebrafish pronephros, brain and Kupffer's vesicle is required for normal organogenesis. *Development* **132**, 1907–1921 (2005).
23. Shu, X. *et al.* Na,K-ATPase  $\alpha 2$  and Ncx4a regulate zebrafish left-right patterning. *Development* **134**, 1921–1930 (2007).
24. Neilson, L. I. *et al.* cDNA cloning and characterization of a human sperm antigen (SPAG6) with homology to the product of the *Chlamydomonas* PF16 locus. *Genomics* **60**, 272–280 (1999).
25. Sapiro, R. *et al.* Sperm antigen 6 is the murine homologue of the *Chlamydomonas reinhardtii* central apparatus protein encoded by the PF16 locus. *Biol. Reprod.* **62**, 511–518 (2000).
26. Zhang, Z. *et al.* A sperm-associated WD repeat protein orthologous to *Chlamydomonas* PF20 associates with Spag6, the mammalian orthologue of *Chlamydomonas* PF16. *Mol. Cell. Biol.* **22**, 7993–8004 (2002).
27. Smith, E. F. & Lefebvre, P. A. PF20 gene product contains WD repeats and localizes to the intermicrotubule bridges in *Chlamydomonas* flagella. *Mol. Biol. Cell* **8**, 455–467 (1997).
28. Dawe, H. R., Farr, H. & Gull, K. Centriole/basal body morphogenesis and migration during ciliogenesis in animal cells. *J. Cell Sci.* **120**, 7–15 (2007).
29. Tay, S. Y., Ingham, P. W. & Roy, S. A homologue of the *Drosophila* kinesin-like protein Costal2 regulates Hedgehog signal transduction in the vertebrate embryo. *Development* **132**, 625–634 (2005).
30. Varjosalo, M., Li, S. P. & Taipale, J. Divergence of hedgehog signal transduction mechanism between *Drosophila* and mammals. *Dev. Cell* **10**, 177–186 (2006).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank H. Bourne, C. C. Hui, Z. Zhang, J. Strauss III and W. Hwang for constructs, antibodies and sharing of unpublished results; R. Harland and T. Mikawa for *X. tropicalis* and chicken tissue; K. Thorn and S. Dandekar for assistance with microscopy and ciliary beat frequency analysis; M.-L. Cheong and Y. Nozawa for technical assistance; and D. Casso, S. Coughlin, T. Kornberg, W. Marshall, T. Mikawa, K. Wemmer and members of the Chen and Chuang laboratories for discussion and critical reading of the manuscript. Some data for this study were acquired at the Nikon Imaging Center at UCSF/QB3. This work was supported by grants from the National Institutes of Health to J.-N.C. and P.-T.C., and a Career Investigator Award from the American Lung Association to P.-T.C.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to P.-T.C. ([pao-tien.chuang@ucsf.edu](mailto:pao-tien.chuang@ucsf.edu)).



## METHODS

**Animal husbandry.** *Fu*<sup>+/-</sup> mice were maintained as described<sup>6</sup>. Wild-type AB fish were used and raised as described<sup>31</sup>. The *smoothened* (also known as *slow muscle omitted*) allele<sup>20</sup> used in this study is *smo*<sup>hi1640Tg</sup>.

**Molecular biology.** Standard molecular biology techniques, including molecular cloning, genomic DNA preparation, RNA isolation, PCR, RT-PCR and Southern analysis were performed as described<sup>32,33</sup>. *Fu*-Flag, *Fu*-4×Flag, *Fu*-mCherry, *Kif7*-3×Myc, *Kif27*-3×Myc, *Kif27*-GFP, SPAG16L-3×HA, and SPAG6-3×HA were cloned into pCAGGS (for immunoprecipitation and immunofluorescence in mammalian cells), pCS2+ (for expression in zebrafish), pcDNA3 (for immunoprecipitation and immunofluorescence in mammalian cells), or FuPw (for lentiviral expression) vectors. Detailed methods and maps are available on request.

FuPw vector (courtesy of K. Wong and H. Bourne) contains the HIV-1 flap sequence, the human polyubiquitin C promoter, a multiple cloning site, and the woodchuck hepatitis virus post-transcriptional regulatory element. Flanking this cassette are 5' and 3' self-inactivating long-terminal repeats. Expression constructs were co-transfected with the HIV packaging vector pCMVΔ8,9 and the envelope glycoprotein vector pVSV-G into HEK293T cells using Lipofectamine 2000 (Invitrogen).

**Morpholino injections.** Wild-type zebrafish embryos were injected with 1.6–4 ng *fu* or 8–12 ng *kif7* or 0.2 ng *ptc1* MO at the one- to two- cell stage. Fluorescein-tagged *fu* morpholino (4 ng) was injected into the yolk of 128-cell-stage embryos to target dorsal forerunner cells. A *p53* morpholino was co-injected with *fu* or *kif7* morpholino at the same concentration to block non-specific cell death<sup>34</sup>. In rescue experiments, 400 pg of mouse *Fu* mRNA was co-injected with *fu* morpholino. In testing genetic epistasis, 0.2 ng of *ptc1* and 2 ng of *fu* MO were co-injected. The *fu* (5'-TGGTACTGATCCATCTCCAGCGACG-3'), *kif7* (5'-GCCGACTCCTTTGGAGACATAGCT-3') and *ptc1* MO (5'-CATAGTCCAAACGGGAGGCAGAGA-3') were described previously<sup>19</sup>.

**In situ hybridization.** Histological analysis and section *in situ* hybridization using <sup>33</sup>P-labelled riboprobes were performed as described<sup>6</sup>. Probes for chick, zebrafish and *X. tropicalis* *Fu* were amplified by PCR using partial or full-length cDNAs (Open Biosystems) as templates. Zebrafish embryos were raised in medium treated with 0.2 mM 1-phenyl-1-2-thiourea to maintain optical transparency. Whole mount *in situ* hybridization was performed as described<sup>35</sup>; probes used were *cmlc2*, *fkdl*, *fkdl*, *nkx2.2b*, *fused*, *shh*, *ptc1*, *spaw* and *pitx2*.

**Ciliary beat frequency and waveform measurements.** Tracheae were dissected out from P10–P14 wild-type and *Fu*<sup>-/-</sup> animals, and cut into rings or strips. Tracheae were washed briefly in PBS and placed in DMEM supplemented with 10% FBS, penicillin–streptomycin and L-glutamate. Tissue was placed in a few drops of medium in a 35-mm glass bottom microwell dish (MatTek). Cilia beating was observed using DIC microscopy on a Nikon TE2000E inverted microscope equipped with Perfect Focus, a ×60 water immersion objective, ×1.5 zoom adaptor and an *in vivo* Scientific incubator set at 37 °C and 5% CO<sub>2</sub>. A Photometrics Coolsnap HQ2 camera and NIS Elements 2.3 software were used to acquire videos of beating cilia at frame rates of 60–70 f.p.s., depending on the size of the defined region of interest (ROI). Ciliary beat frequency was measured by defining an ROI in the upper third of the ciliary shaft, and plotting the changes in pixel intensity over time in the obtained image series. This data was subsequently Fourier transformed to obtain the frequency using MatLab. Waveform was analysed by tracing of cilia from individual movie frames in Adobe Illustrator, or by manual tracking using the MtrackJ plugin (E. Meijering, Biomedical Imaging Group, University Medical Center, Rotterdam) in NIH ImageJ.

**Cell culture, transfections and immunoprecipitation.** HEK 293T cells were maintained in DMEM supplemented with 10% FBS, penicillin–streptomycin and L-glutamate. Cells were transfected with Lipofectamine 2000 (Invitrogen)

according to manufacturer's instructions. Forty-eight hours after transfection, cells were collected and lysed in lysis buffer (1% Triton X-100, 150 mM NaCl, 50 mM Tris-HCl, pH 7.5, 1 mM EDTA, 0.5 mM PMSF, 2 μg ml<sup>-1</sup> pepstatin A, 10 μg ml<sup>-1</sup> leupeptin, 5 μg ml<sup>-1</sup> aprotinin). Lysates were sheared with a 20-gauge needle and remained on ice for 30 min. Lysates were then clarified by centrifugation at 20,817g for 20 min at 4 °C. The supernatant was removed and bound to 50 μl of anti-Flag M2 agarose beads (Sigma) for 4 h at 4 °C with constant nutation. Beads were washed five times with lysis buffer before the addition of sample buffer. Immunoprecipitated proteins were analysed by 7.5% SDS-PAGE and transferred to PVDF for immunoblotting. Antibodies used were rabbit anti-Flag (Sigma, 1:2,000), rabbit anti-Myc (Sigma, 1:2,000), and rabbit anti-HA (Sigma, 1:1,000).

**Primary MTEC culture and viral transduction.** Primary MTECs were derived from P10–P21 mice and cultured as described<sup>36</sup>. Lentivirus was produced by co-transfecting cDNAs cloned into the FuPw vector with pCMVΔ8,9 and pVSV-G into HEK 293T cells as described above. Supernatant was collected 72–96 h after transfection, filtered through a 0.45 μm PES membrane syringe filter unit (Nalgene), and concentrated tenfold using a Centrprep Ultracel YM-10 device (Millipore). Infection of MTECs was performed as described<sup>37</sup>.

**Immunofluorescence and microscopy.** Cells were fixed in 4% paraformaldehyde for most applications, or in ice-cold methanol for visualization of basal bodies. Standard procedures were used for immunostaining. Primary antibodies used were mouse anti-acetylated-α-tubulin (Sigma, 1:2,000) and mouse anti-γ tubulin (Sigma, 1:2,000). Secondary antibodies and conjugates used were donkey anti-mouse AlexaFluor 594 (Molecular Probes, 1:2,000), donkey anti-mouse FITC (Molecular Probes, 1:2,000), and rhodamine-conjugated phalloidin (Sigma, 1:200). Fluorescent confocal images were acquired using a Nikon TE2000U inverted microscope with a Yokogawa CSU22 spinning disk confocal (Solamere Technology Group), a Photometrics Cascade II Camera, and MicroManager software (Vale laboratory, University of California–San Francisco). Images were acquired with a ×100 oil-immersion lens and a ×1.5 zoom adaptor (Nikon) using two laser lines (488 nm and 568 nm). Confocal stacks were collected using a 0.25-μm step size along the z-axis. Stacks were analysed and xy, xz, and yz projections were generated using ImageJ and the VolumeViewer plugin (K. U. Barthel, Internationale Medieninformatik). Deconvolution was performed with the Iterative Deconvolve 3D plugin (R. Dougherty, OptiNav, Inc.).

**Immunohistochemistry staining.** Immunohistochemistry staining using anti-Engrailed (4D9, Developmental Studies Hybridoma Bank) at 1:100 dilution and anti-acetylated tubulin (Sigma) at 1:200 dilution was conducted as described<sup>23</sup>. Confocal images were acquired with an LSM510 confocal microscope (Zeiss).

**Fluorescent bead injection.** Fluorescent beads diluted 1:100 in PBS were injected into Kupffer's vesicle at the 8–10-somite stage<sup>23</sup>. Embryos were imaged on a Zeiss Axioplan 2 microscope using a ×63 water immersion lens (Zeiss).

31. Westerfield, M. *The Zebrafish Book* (Univ. Oregon Press, 1995).

32. Nagy, A., Gertsenstein, M., Vintersten, K. & Behringer, R. *Manipulating the Mouse Embryo: A Laboratory Manual* 3rd edn (Cold Spring Harbour Laboratory Press, 2003).

33. Sambrook, J. & Russell, D. W. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbour Laboratory Press, 2001).

34. Eisen, J. S. & Smith, J. C. Controlling morpholino experiments: don't stop making antisense. *Development* **135**, 1735–1743 (2008).

35. Chen, J. N. & Fishman, M. C. Zebrafish tinman homolog demarcates the heart field and initiates myocardial differentiation. *Development* **122**, 3809–3816 (1996).

36. You, Y., Richer, E. J., Huang, T. & Brody, S. L. Growth and differentiation of mouse tracheal epithelial cells: selection of a proliferative population. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **283**, L1315–L1321 (2002).

37. Vladar, E. K. & Stearns, T. Molecular characterization of centriole assembly in ciliated epithelial cells. *J. Cell Biol.* **178**, 31–42 (2007).

# Haematopoietic stem cells depend on $G\alpha_s$ -mediated signalling to engraft bone marrow

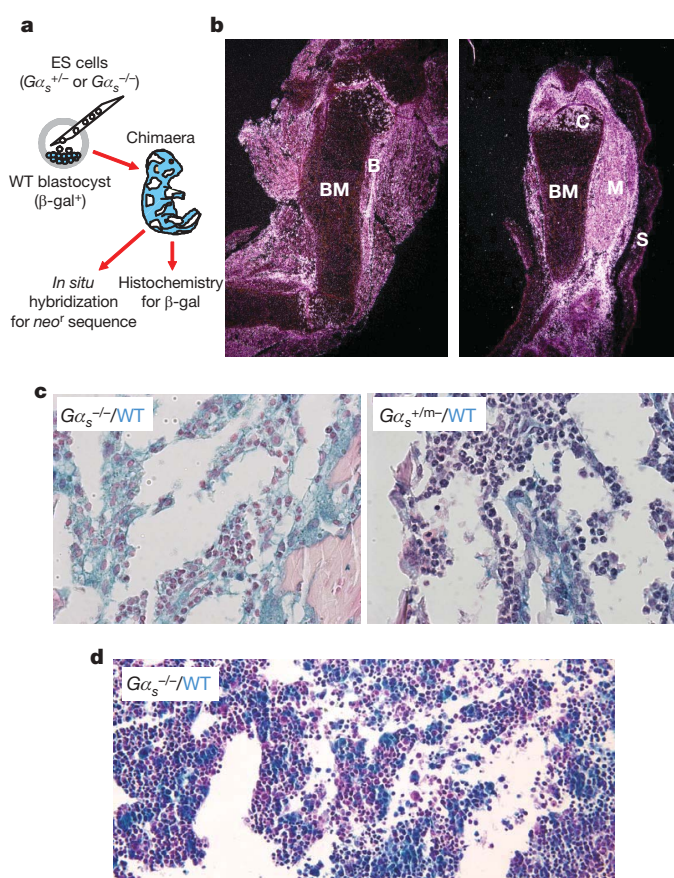
Gregor B. Adams<sup>1,4†</sup>, Ian R. Alley<sup>1</sup>, Ung-il Chung<sup>2†</sup>, Karissa T. Chabner<sup>1</sup>, Nathaniel T. Jeanson<sup>1</sup>, Cristina Lo Celso<sup>1,4</sup>, Emily S. Marsters<sup>1</sup>, Min Chen<sup>6</sup>, Lee S. Weinstein<sup>6</sup>, Charles P. Lin<sup>3</sup>, Henry M. Kronenberg<sup>2</sup> & David T. Scadden<sup>1,4,5</sup>

Haematopoietic stem and progenitor cells (HSPCs) change location during development<sup>1</sup> and circulate in mammals throughout life<sup>2</sup>, moving into and out of the bloodstream to engage bone marrow niches in sequential steps of homing, engraftment and retention<sup>3–5</sup>. Here we show that HSPC engraftment of bone marrow in fetal development is dependent on the guanine-nucleotide-binding protein stimulatory  $\alpha$  subunit ( $G\alpha_s$ ). HSPCs from adult mice deficient in  $G\alpha_s$  ( $G\alpha_s^{-/-}$ ) differentiate and undergo chemotaxis, but also do not home to or engraft in the bone marrow in adult mice and demonstrate a marked inability to engage the marrow microvasculature. If deleted after engraftment,  $G\alpha_s$  deficiency did not lead to lack of retention in the marrow, rather cytokine-induced mobilization into the blood was impaired. Testing whether activation of  $G\alpha_s$  affects HSPCs, pharmacological activators enhanced homing and engraftment *in vivo*.  $G\alpha_s$  governs specific aspects of HSPC localization under physiological conditions *in vivo* and may be pharmacologically targeted to improve transplantation efficiency.

$G\alpha_s$  is essential for normal development, with gene-deficient embryos showing fatal defects before day E10.5 (ref. 6). To overcome this limitation, chimaeric mice were generated by blastocyst injection of  $G\alpha_s^{-/-}$  embryonic stem cells<sup>7</sup> into a blastocyst transgenic for  $\beta$ -galactosidase ( $\beta$ -gal), enabling resultant embryos to be assessed for tissue chimaerism using *in situ* hybridization against *neo* in the targeting vector to score for  $G\alpha_s^{-/-}$  (knockout) cells and histochemical staining for  $\beta$ -gal to score for  $G\alpha_s^{+/+}$  (wild-type) cells (Fig. 1a). On examining E17.5 embryos by *in situ* hybridization, chimaerism could be readily detected. Limb sections revealed *neo*-expressing  $G\alpha_s^{-/-}$  cell involvement in skin, muscle, bone and cartilage, but without evidence of contribution to bone marrow (Fig. 1b). In contrast, as a positive control for the *in situ* hybridization, we examined limb sections from mice chimaeric for an unrelated knockout and could readily detect bone-marrow chimaerism (Supplementary Fig. 1). Independent analysis using  $\beta$ -gal staining for wild-type cells revealed only  $\beta$ -gal<sup>+</sup> cells in the bone marrow of  $G\alpha_s^{-/-}$  chimaeras, but both  $\beta$ -gal<sup>+</sup> and  $\beta$ -gal<sup>-</sup> cells in the bone marrow of mice heterozygous for deletion of the maternal copy of  $G\alpha_s$  ( $G\alpha_s^{+/m-}$ ) (100% versus 44%  $\beta$ -gal<sup>+</sup>, respectively; Fig. 1c). Other tissues of  $G\alpha_s^{-/-}$  chimaeras demonstrated both  $\beta$ -gal<sup>+</sup> and  $\beta$ -gal<sup>-</sup> cells (Supplementary Fig. 2). Therefore, the bone marrow appeared distinctive in its lack of contribution from  $G\alpha_s^{-/-}$  cells.

Evaluation of E13.5 fetal liver demonstrated significant chimaerism with  $G\alpha_s^{-/-}$  cells by  $\beta$ -gal (48%  $\beta$ -gal<sup>+</sup>; Fig. 1d) and no difference in colony-forming unit-culture (CFU-C) content between  $G\alpha_s^{+/+}$  and  $G\alpha_s^{-/-}$  chimaeric mononuclear cells (MNCs; Supplementary Fig. 3). Therefore, HSPCs form and have intact haematopoietic potential in

the fetal liver even in the absence of  $G\alpha_s$ , but without  $G\alpha_s$ , there is a failure to engraft bone marrow by E17.



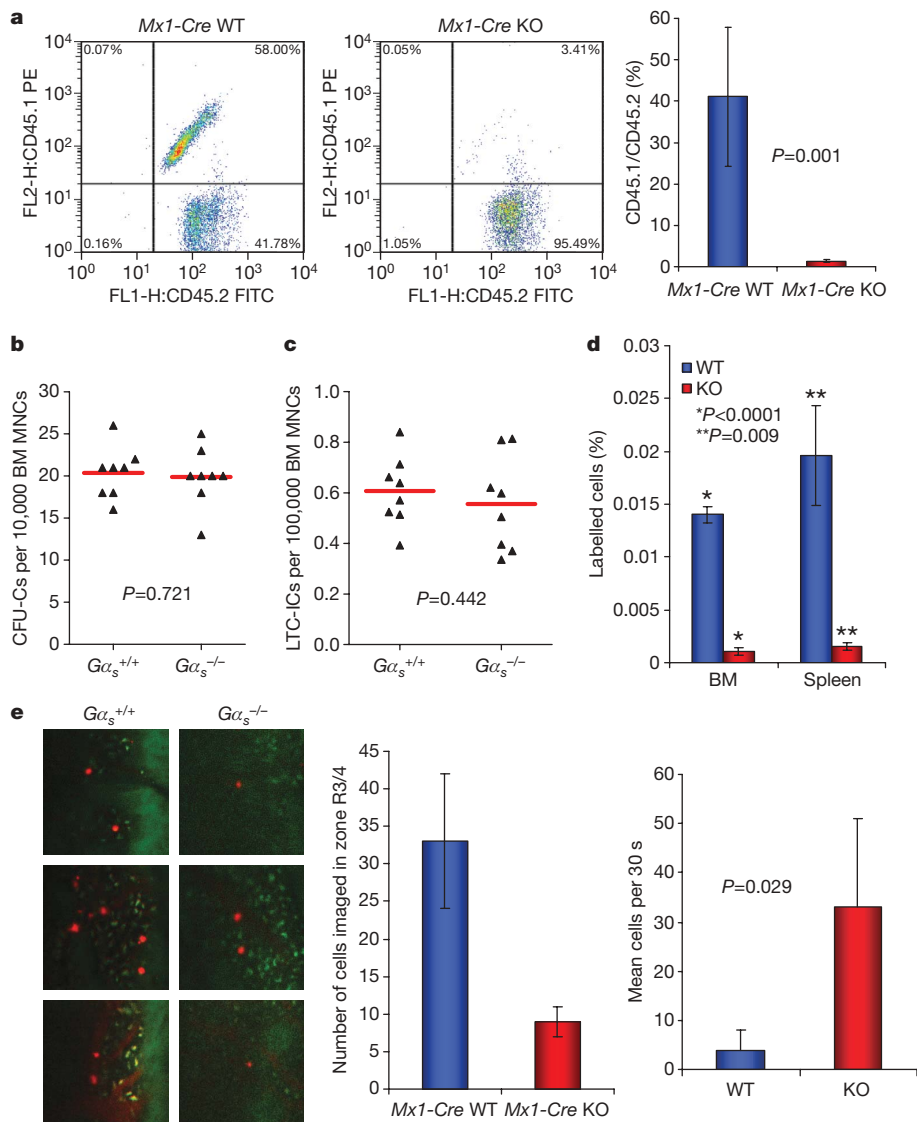
**Figure 1 |  $G\alpha_s$  is required for HSPC engraftment of bone marrow in development.** **a**, Chimaeric mice were created by injection of  $G\alpha_s^{-/-}$  or  $G\alpha_s^{+/+}$  embryonic stem cells into wild-type or  $\beta$ -galactosidase transgenic blastocysts. At E17.5, the mice were killed and the organs were probed for  $G\alpha_s^{-/-}$  cells (using a *neo* probe). **b**, *In situ* hybridization of femur and tibia is shown. Bone marrow (BM), bone (B), muscle (M), skin (S) and chondrocytes (C) are indicated. Original magnification,  $\times 40$ . **c**, Mice were assessed by histochemistry for  $\beta$ -gal (blue stain) in  $G\alpha_s^{-/-}$  (left panel) or  $G\alpha_s^{+/m-}$  (right panel) chimaeric animal bone marrow. Original magnification,  $\times 200$ . **d**, Histochemical analysis of  $\beta$ -gal expression in fetal liver from E13.5  $G\alpha_s^{-/-}$  chimaeric mice. Original magnification,  $\times 100$ .

<sup>1</sup>Center for Regenerative Medicine, <sup>2</sup>Endocrine Unit and <sup>3</sup>Advanced Microscopy Program, Center for Systems Biology and Wellman Center for Photomedicine, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts 02114, USA. <sup>4</sup>Harvard Stem Cell Institute, <sup>5</sup>Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, Massachusetts 02138, USA. <sup>6</sup>National Institute for Diabetes, Digestive and Kidney Diseases, Bethesda, Maryland 20892, USA. <sup>†</sup>Present addresses: Eli and Edythe Broad Center for Regenerative Medicine and Stem Cell Research at USC, Keck School of Medicine, University of Southern California, Los Angeles, California 90033, USA (G.B.A.); Department of Bioengineering, Graduate Schools of Engineering and Medicine, University of Tokyo, Bunkyo-ku, Tokyo 113-0033, Japan (U.-i.C.).

To evaluate whether this developmental requirement was also present in adult HSPCs, we crossed  $G\alpha_s^{\Delta/\Delta}$  (exon 1 of the  $G\alpha_s$  gene flanked by *loxP* sites) mice with *Mx1-Cre* mice, a strain that enables high-efficiency polyI:C-induced Cre expression in HSPCs. Competitive repopulation was examined using bone-marrow MNCs obtained from 6-week-old  $G\alpha_s^{+/+}$  *Mx1-Cre*<sup>+</sup> or  $G\alpha_s^{\Delta/\Delta}$  *Mx1-Cre*<sup>+</sup> mice treated with polyI:C for 5 days (hereafter referred to as *Mx1-Cre* wild type and *Mx1-Cre* knockout, respectively), and transplanted in competition (1:1) with wild-type cells. Deletion of the  $G\alpha_s$  gene in primitive cells was confirmed by polymerase chain reaction (PCR; Supplementary Fig. 4). Analysis of the number of HSPCs in the bone-marrow MNCs, phenotypically defined as  $\text{Lin}^- \text{c-Kit}^+ \text{Sca1}^+ \text{Flk2}^-$ , was also identical between the *Mx1-Cre* wild-type and *Mx1-Cre* knockout animals (wild type,  $0.031 \pm 0.012\%$ ; knockout,  $0.038 \pm 0.002\%$ ;  $P = 0.545$ ). Cells from the *Mx1-Cre* wild-type animals were able to engraft the bone marrow efficiently, whereas bone-marrow MNCs from the *Mx1-Cre* knockout mice were essentially absent from the bone marrow at early (4 weeks)

and late (16 weeks) time points after transplantation (Fig. 2a). Furthermore, when animals were transplanted with bone marrow cells from *Mx1-Cre* knockout mice in the absence of competing wild-type cells, spleen engraftment was significantly reduced as assessed by day 8 colony-forming unit-spleen (CFU-S<sub>8</sub>) number (Supplementary Fig. 5), and all animals died within 2 weeks, indicating a failure of extramedullary engraftment.  $G\alpha_s$  is therefore required for HSPC engraftment in adult as well as in embryonic settings.

A failure to engraft successfully may reflect multiple abnormalities, which we sequentially examined. As others had previously found that pharmacological modifiers of the adenylyl cyclase pathways influenced myeloid differentiation *in vitro*<sup>8,9</sup>, bone-marrow MNCs were obtained from *Mx1-Cre* wild-type and knockout mice and progenitor cell activity was analysed using the CFU-C assay. No significant differences were observed between the genotypes in colony number (Fig. 2b), morphology or size. Similarly, analysis of the more primitive cell fraction using the long-term culture initiating cell (LTC-IC) assay<sup>10</sup>



**Figure 2 |  $G\alpha_s$  is required for HSPC engraftment of bone marrow in adults.**  $G\alpha_s$  was conditionally deleted in haematopoietic cells of 6-week-old  $G\alpha_s^{\Delta/\Delta}$  *Mx1-Cre* mice and cells were then assessed for engraftment capability. **a**, Representative flow cytometric analysis at 16 weeks and collated data of engraftment of *Mx1-Cre* wild-type ( $G\alpha_s^{+/+}$ ) or *Mx1-Cre* knockout ( $G\alpha_s^{-/-}$ ) cells in competitive repopulation assay ( $n = 8$  from 2 independent experiments; error bars represent s.e.m.). **b**, **c**, Bone-marrow (BM) MNCs were assessed for their *in vitro* growth potential using CFU-C (**b**) and LTC-IC (**c**) assays. **d**, *In vivo* homing to wild-type bone marrow and spleen of

*Mx1-Cre* wild-type and knockout LKS<sup>+</sup> cells ( $n = 3$ ; error bars represent s.e.m.). **e**, Intravital imaging of calvarium bone marrow of wild-type recipient mice injected via the tail vein with *Mx1-Cre* wild-type or knockout LKS<sup>+</sup> cells stained with Vybrant DiD (1,1'-dioctadecyl-3,3,3'-tetramethylindodicarbocyanine perchlorate). Representative images (red, DiD; green, autofluorescence; field of view,  $330 \times 330 \mu\text{m}$ ; original magnification,  $\times 30$ ) and quantification (graphs) of the number of cells visualized over a 30-s interval at  $\sim 35$  min after infusion to be stably localized or circulating is shown ( $n = 3$ ; error bars represent s.d.).

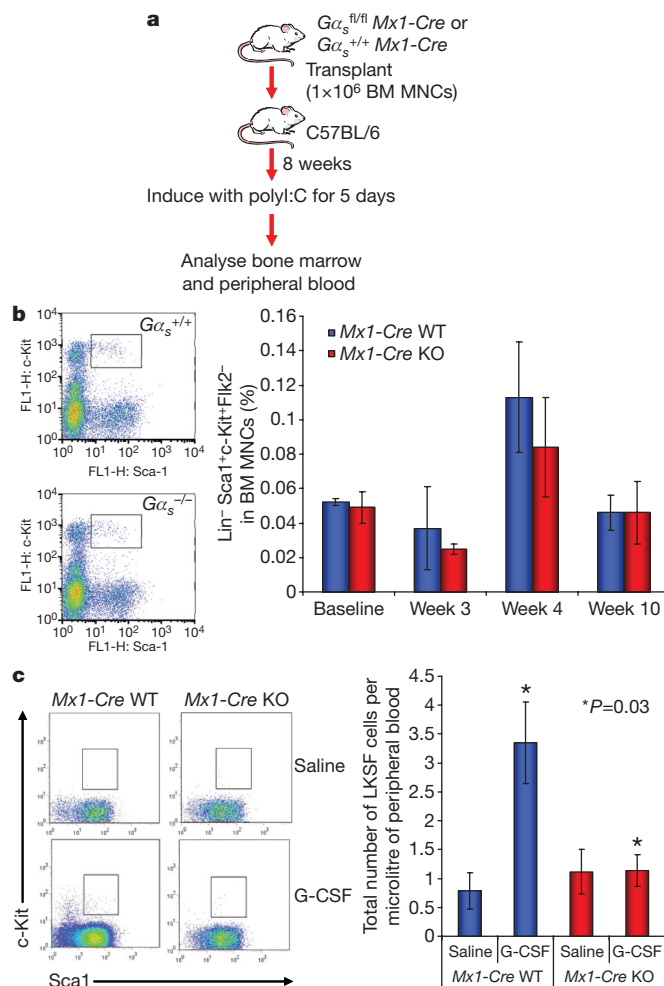


demonstrated no significant differences between the wild-type and knockout cells (Fig. 2c). Taken together, these data demonstrate that deletion of the  $G\alpha_s$  gene does not alter the differentiation potential of primitive bone-marrow MNCs *in vitro*.

Migratory capacity was next assessed using the primitive Lin<sup>-</sup>c-Kit<sup>+</sup>Sca1<sup>+</sup> (LKS<sup>+</sup>) or more mature Lin<sup>-</sup>c-Kit<sup>+</sup>Sca1<sup>-</sup> (LKS<sup>-</sup>) bone-marrow MNCs of *Mx1-Cre* wild-type or *Mx1-Cre* knockout mice. *In vitro* chemotaxis towards SDF-1 $\alpha$ , the only known *in vitro* chemotactic agent for HSPCs<sup>11</sup>, demonstrated comparable migratory capacity for each genotype of LKS<sup>+</sup> cells for all but the lowest concentration of SDF-1 $\alpha$  (Supplementary Fig. 6a). In addition, there was no significant change in the chemotaxis of the more mature LKS<sup>-</sup> subset of cells (Supplementary Fig. 6b) or lymph node lymphocytes to SDF-1 $\alpha$ , (data not shown) indicating that there was no generalized impairment of *in vitro* cell motility related to the  $G\alpha_s$  defect.

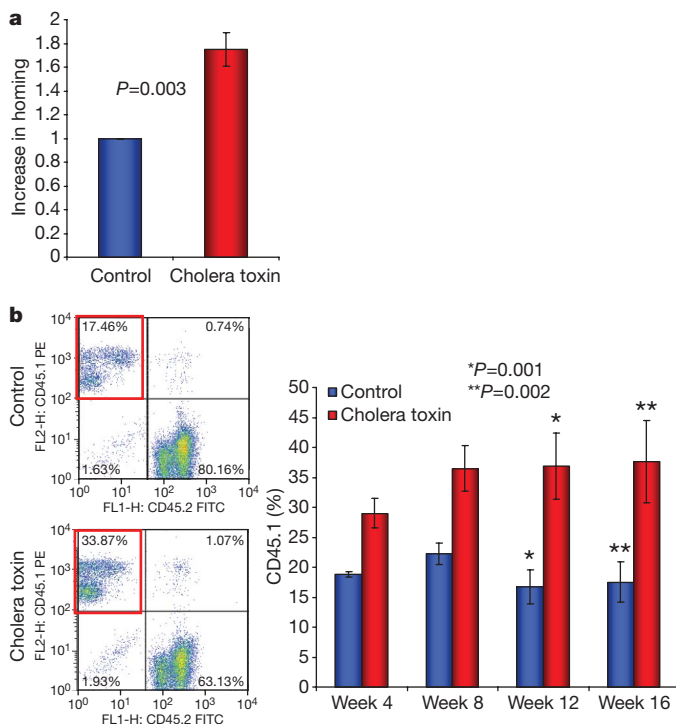
Primitive cells isolated from bone-marrow MNCs of *Mx1-Cre* wild-type or knockout mice were then evaluated for *in vivo* capacity to home to the bone-marrow environment. Fluorescently labelled LKS<sup>+</sup> cells were injected into the peripheral circulation of mice and the numbers of cells present in the bone marrow and spleen were measured using flow cytometry at 6 h after injection. Deletion of  $G\alpha_s$  in LKS<sup>+</sup> bone-marrow MNCs resulted in a marked impairment in their ability to home to both the bone marrow and spleen (Fig. 2d). To examine whether this was specific to primitive cell homing to the bone marrow, or a more general defect, we examined the ability of the more mature LKS<sup>-</sup> and lymph node lymphocytes to home to haematopoietic organs. Whereas LKS<sup>-</sup> cells demonstrated a similar impairment as LKS<sup>+</sup> cells (Supplementary Fig. 7a), no defect in  $G\alpha_s$ <sup>-/-</sup> lymphocytes (genotype confirmed by PCR) homing to lymph node, bone marrow or spleen was observed (Supplementary Fig. 7b). These data argue against a generalized defect in cell homing associated with a deficiency of  $G\alpha_s$  and identify a specific bone-marrow homing defect of primitive haematopoietic cells. To assess the specific homing defect of primitive  $G\alpha_s$ <sup>-/-</sup> cells, we visualized HSPC homing in the calvarial bone-marrow endothelium using intravital confocal/two-photon videomicroscopy<sup>12</sup>. We found that  $G\alpha_s$ <sup>-/-</sup> LKS<sup>+</sup> cells were defective in their interaction with the bone-marrow endothelium with significantly fewer cells observed rolling on the endothelial cells, whereas the number of cells retained in the circulation was significantly increased (Fig. 2e and Supplementary Movies 1 and 2). Therefore,  $G\alpha_s$  is critical for the engagement of the bone-marrow vasculature by primitive haematopoietic cells, thereby enabling subsequent steps in stem-cell engraftment.

Deletion of the  $G\alpha_s$  subunit resulted in significant reductions in the ability of primitive cells to establish haematopoiesis in the bone-marrow environment. Therefore, we examined whether the same deletion had any effects once engraftment of the cells in the bone marrow had been achieved. To study this, bone-marrow MNCs from *Mx1-Cre* wild-type or *Mx1-Cre* knockout mice were transplanted into wild-type mice. Eight weeks after transplantation, the mice were treated with polyI:C to induce gene deletion (Fig. 3a). Longitudinal analysis demonstrated no significant differences between mice with *Mx1-Cre* wild-type or knockout cells in terms of peripheral blood complete blood count (CBC), bone-marrow CFU-C frequency (Supplementary Fig. 8a–c) or HSPC frequency as assessed by immunophenotype (Fig. 3b). Specific analysis of HSPC cell cycling status or number of apoptotic cells also indicated that there were no differences between the two different cell genotypes (Supplementary Fig. 9a, b). Furthermore, there was no evidence of HSPC enriched cell movement from the bone marrow to the peripheral blood by immunophenotypic analysis to indicate a retention defect. Rather, we did observe a striking defect in the ability of the *Mx1-Cre* knockout cells to be mobilized using granulocyte colony-stimulating factor (G-CSF; Fig. 3c). These data demonstrate that  $G\alpha_s$  signalling is not required for HSPC retention in the marrow, but under conditions of stress, mimicked here by G-CSF stimulation, trafficking out of the marrow space is impaired.



**Figure 3 |  $G\alpha_s$  signalling is not required for retention of the HSPCs in the bone marrow, but does influence mobilization by G-CSF.** **a**, Diagrammatic representation of the procedure to assess  $G\alpha_s$  signalling in HSPC bone-marrow retention.  $G\alpha_s^{+/-}$  *Mx1-Cre* or  $G\alpha_s^{fl/fl}$  *Mx1-Cre* cells were injected into wild-type mice; after 8 weeks, deletion of the  $G\alpha_s$  gene was induced by polyI:C and bone marrow and peripheral blood evaluated. **b**, Retention of primitive LKS Flk2<sup>-</sup> *Mx1-Cre* wild-type or knockout cells in the bone marrow. Representative flow cytometric analysis at week 10 and mean collated data are shown ( $n = 5$ ; error bars represent s.e.m.). **c**, Representative flow cytometric plots (left) and quantification (right) of HSPC mobilization into the peripheral circulation after 5 days of G-CSF. Sca1 and c-Kit expression of gated Lin<sup>-</sup>Flk2<sup>-</sup> cells is shown ( $n = 5$ ; error bars represent s.e.m.).

Genetic alteration of  $G\alpha_s$  signalling demonstrated a central role in bone marrow homing of HSPCs. However, to determine if these findings might be relevant for medicine, we tested whether pharmacological modifiers of  $G\alpha_s$  signalling could alter stem-cell function *in vivo*. We treated bone-marrow MNCs from wild-type C57BL/6 mice with cholera toxin, a compound known to constitutively activate  $G\alpha_s$  by preventing GTP hydrolysis from the ADP-ribose- $G\alpha_s$ -GTP complex. Using a simple *ex vivo* exposure of cells for 1 h, we observed increased intracellular cAMP concentrations within haematopoietic cells by enzyme-linked immunosorbent assay (ELISA; data not shown). This resulted in an enhancement in the ability of treated cells to both home and engraft in the bone marrow, with approximately twofold increased engraftment at 16 weeks (Fig. 4a, b). To test whether the enhancement by cholera toxin was due to a nonspecific action of the compound, homing of  $G\alpha_s$ <sup>-/-</sup> cells treated with cholera toxin was assessed; however, no improvement in the homing defect was observed (Supplementary Fig. 10). Because stimulation of  $G\alpha_s$ -coupled receptors has also been noted to increase



**Figure 4 | Pharmacological modulation of  $G\alpha_s$  affects homing and engraftment of primitive wild-type bone-marrow MNCs.** **a**, Primitive Lin<sup>−</sup> bone-marrow MNCs pre-treated with cholera toxin or mock treated for 1 h *ex vivo* were assessed for their *in vivo* homing potential ( $n = 3$  from 3 individual experiments; error bars represent s.e.m.). **b**, Engraftment potential in a competitive transplant model ( $n = 9$  from 2 individual experiments; error bars represent s.e.m.).

proliferation of HSPCs in developing zebrafish<sup>13</sup>, we tested whether murine LKS<sup>+</sup> cells treated with cholera toxin demonstrated increased cell cycling after transplantation. Cell cycle evaluation of labelled cells 2 days after injection into lethally irradiated hosts did not detect any difference between untreated and cholera-toxin-treated cells (Supplementary Fig. 11). Therefore, cholera toxin acting at least in part through the  $G\alpha_s$ -coupled pathway augments the homing of the cells and engraftment in the bone marrow. This effect on engraftment was durable and without distortion of mature cell subsets, indicating that the cholera-toxin-induced improvement in HSPC engagement of the niche did not compromise cellular function.

Elucidating the mechanisms by which HSPCs home to the bone-marrow environment has biological and medical implications that have made it an area of intense investigation. A number of molecules have been implicated; often they were studied because of their effects on other haematopoietic populations such as lymphocytes. Here we show that  $G\alpha_s$  has a key role in the specific homing of HSPCs to the bone marrow, in a manner not shared by lymphocytes.

The inability of stem cells to migrate to the bone marrow in development strongly resembles the phenotype observed with the deletion of CXCR4 or SDF-1 $\alpha$  (refs 14, 15). This, coupled with an impairment in SDF-1 $\alpha$ -responsive migration at specific concentrations *in vitro*, suggests that CXCR4 may be using  $G\alpha_s$  as its predominant signalling pathway in that cell type. The data presented here do not rule out this possibility for fetal haematopoiesis. However, two important distinctions between the  $G\alpha_s$  and CXCR4 null phenotypes exist for stem-cell function in adult animals. First, pharmacological interruption of CXCR4 signalling results in prompt mobilization of stem cells into the circulation<sup>16</sup>. Genetic disruption of  $G\alpha_s$  resulted in no such change in location. Indeed, G-CSF-induced mobilization of cells from the bone marrow was impaired in the absence of  $G\alpha_s$ , indicating that egress from the marrow may be diminished. Therefore, whereas CXCR4 is a key component of HSPC retention in the bone marrow<sup>17</sup>,

it is not exerting its effect through  $G\alpha_s$ . Second, CXCR4 null cells engraft in the bone marrow, whereas  $G\alpha_s$  null cells do not<sup>18</sup>. It is therefore unlikely that the role of  $G\alpha_s$  is simply downstream of CXCR4 as this molecule is dispensable for HSPC engraftment in the transplant setting.

Our results demonstrate that  $G\alpha_s$  is critical at the HSPC homing stage, enabling engagement of the niche; however, continued  $G\alpha_s$  signalling is not required to retain the cells there. There is then a meaningful molecular distinction between the homing, engraftment and retention processes. It could be hypothesized that the deficiency in the ability of the  $G\alpha_s^{-/-}$  cells to home to the bone marrow could lead to a depletion of the HSPCs in the marrow space and an accumulation in the blood. However, the decreased egress from the marrow may balance the decreased ingress. Alternatively, there may be compensatory changes in resident cells that mask changes in the population.

Another unknown from these studies is the potential upstream receptor using this pathway. The candidates are numerous and may include adrenergic and prostaglandin receptors. The report of prostaglandin E<sub>2</sub> (PGE<sub>2</sub>) serving as a means of enhancing stem cell transplantation in mice<sup>13</sup> indicates that this is probably at least one receptor capable of activating the pathway. Whether PGE<sub>2</sub> mediates an effect physiologically in mammals could not be discerned from the data of ref. 13; however, our data indicate that the downstream pathway is critical for physiological processes in development.

The localization of stem cells after transplantation is a critical determinant of success of that clinical procedure. Currently, massive numbers of stem cells are required in clinical transplantation in part due to the limited efficiency of homing and engraftment. This is particularly problematic in umbilical cord blood transplantation where the number of stem cells is limited. This problem is being approached by dual unit transplantation, an expensive undertaking that indicates how a modest (twofold) increase in stem cells can profoundly affect clinical outcomes. Alternatives to the use of multiple unit infusions might be to increase the homing efficiency of the cells, increase the nurturing capacity of the niche or increase the number of cells by *ex vivo* expansion. Each approach has potential and with new information about the means of affecting each parameter, the possibility of translation to clinical trial. Transient exposure to agents stimulating  $G\alpha_s$  is one such candidate approach.

## METHODS SUMMARY

**$G\alpha_s^{-/-}$  chimaeric mice.** Mice were created as described previously<sup>7</sup>.  $G\alpha_s^{-/-}$  embryonic stem cells were injected into a blastocyst transgenic for  $\beta$ -galactosidase. Resultant embryos were then assessed for tissue chimaerism using *in situ* hybridization against *neo* to detect  $G\alpha_s^{-/-}$  cells or histochemical staining for  $\beta$ -galactosidase to detect  $G\alpha_s^{+/+}$  cells.

**$G\alpha_s$  conditional knockout adult mice.**  $G\alpha_s^{fl/+}$  mice were crossed with Mx1-Cre<sup>+</sup> mice to create  $G\alpha_s^{fl/+}$  Mx1-Cre<sup>+</sup> mice. These mice were then bred to obtain  $G\alpha_s^{fl/fl}$  Mx1-Cre<sup>+</sup> mice and  $G\alpha_s^{+/+}$  Mx1-Cre<sup>+</sup> littermates. Deletion of the  $G\alpha_s$  gene was achieved after exposure of the mice to polyI:C for 5 days. Analysis of the primitive bone-marrow MNCs used standard *in vitro* immunophenotyping, CFU-C, LTC-IC and chemotaxis assays, or *in vivo* homing and engraftment assays.

**$G\alpha_s$  deletion after engraftment.** Bone-marrow MNCs from  $G\alpha_s^{+/+}$  Mx1-Cre<sup>+</sup> or  $G\alpha_s^{fl/fl}$  Mx1-Cre<sup>+</sup> mice were transplanted into wild-type mice. Eight weeks after transplantation, the mice were treated with polyI:C to induce deletion of the gene. Peripheral blood and bone-marrow MNCs were then evaluated at various time points after deletion.

**Cholera toxin treatment.** Bone marrow MNCs from wild-type C57BL/6 mice were treated with cholera toxin (10  $\mu$ g ml<sup>−1</sup>) for 1 h *ex vivo*. Treated cells were then used in *in vivo* homing and engraftment studies.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 23 May 2008; accepted 19 January 2009.

Published online 25 March 2009.

1. Mikkola, H. K. A. & Orkin, S. H. The journey of developing hematopoietic stem cells. *Development* 133, 3733–3744 (2006).

2. Wright, D. E., Wagers, A. J., Gulati, A. P., Johnson, F. L. & Weissman, I. L. Physiological migration of hematopoietic stem and progenitor cells. *Science* **294**, 1933–1936 (2001).
3. Mazo, I. B. *et al.* Hematopoietic progenitor cell rolling in bone marrow microvessels: parallel contributions by endothelial selectins and vascular cell adhesion molecule 1. *J. Exp. Med.* **188**, 465–474 (1998).
4. Papayannopoulou, T., Priestley, G. V., Nakamoto, B., Zafiroopoulos, V. & Scott, L. M. Molecular pathways in bone marrow homing: dominant role of  $\alpha_4\beta_1$  over  $\beta_2$ -integrins and selectins. *Blood* **98**, 2403–2411 (2001).
5. Lapidot, T., Dar, A. & Kollet, O. How do stem cells find their way home? *Blood* **106**, 1901–1910 (2005).
6. Yu, S. *et al.* Variable and tissue-specific hormone resistance in heterotrimeric  $G_s$  protein  $\alpha$ -subunit ( $G_{s\alpha}$ ) knockout mice is due to tissue-specific imprinting of the  $G_{s\alpha}$  gene. *Proc. Natl Acad. Sci. USA* **95**, 8715–8720 (1998).
7. Bastepe, M. *et al.* Stimulatory G protein directly regulates hypertrophic differentiation of growth plate cartilage *in vivo*. *Proc. Natl Acad. Sci. USA* **101**, 14794–14799 (2004).
8. Dexter, T. M., Whetton, A. D. & Heyworth, C. M. Inhibitors of cholera toxin-induced adenosine diphosphate ribosylation of membrane-associated proteins block stem cell differentiation. *Blood* **65**, 1544–1548 (1985).
9. Long, M. W., Heffner, C. H. & Gragowski, L. L. Cholera toxin and phorbol diesters synergistically modulate murine hematopoietic progenitor cell proliferation. *Exp. Hematol.* **16**, 195–200 (1988).
10. Ploemacher, R. E., van der Sluijs, J. P., van Beurden, C. A., Baert, M. R. & Chan, P. L. Use of limiting-dilution type long-term marrow cultures in frequency analysis of marrow-repopulating and spleen colony-forming hematopoietic stem cells in the mouse. *Blood* **78**, 2527–2533 (1991).
11. Wright, D. E., Bowman, E. P., Wagers, A. J., Butcher, E. C. & Weissman, I. L. Hematopoietic stem cells are uniquely selective in their migratory response to chemokines. *J. Exp. Med.* **195**, 1145–1154 (2002).
12. Sipkins, D. A. *et al.* *In vivo* imaging of specialized bone marrow endothelial microdomains for tumour engraftment. *Nature* **435**, 969–973 (2005).
13. North, T. E. *et al.* Prostaglandin E2 regulates vertebrate haematopoietic stem cell homeostasis. *Nature* **447**, 1007–1011 (2007).
14. Nagasawa, T. *et al.* Defects of B-cell lymphopoiesis and bone-marrow myelopoiesis in mice lacking the CXC chemokine PBSF/SDF-1. *Nature* **382**, 635–638 (1996).
15. Ma, Q. *et al.* Impaired B-lymphopoiesis, myelopoiesis, and derailed cerebellar neuron migration in CXCR4- and SDF-1-deficient mice. *Proc. Natl Acad. Sci. USA* **95**, 9448–9453 (1998).
16. Broxmeyer, H. E. *et al.* AMD3100 and CD26 modulate mobilization, engraftment, and survival of hematopoietic stem and progenitor cells mediated by the SDF-1/CXCL12-CXCR4 axis. *Ann. NY Acad. Sci.* **1106**, 1–19 (2007).
17. Foudi, A. *et al.* Reduced retention of radioprotective hematopoietic cells within the bone marrow microenvironment in CXCR4<sup>-/-</sup> chimeric mice. *Blood* **107**, 2243–2251 (2006).
18. Kawabata, K. *et al.* A cell-autonomous requirement for CXCR4 in long-term lymphoid and myeloid reconstitution. *Proc. Natl Acad. Sci. USA* **96**, 5663–5667 (1999).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** Financial support for this work was provided by the Burroughs Wellcome Fund, Doris Duke Charitable Trust (D.T.S.), the Harvard Stem Cell Institute (C.P.L.) and the National Institutes of Health (G.B.A., C.P.L., H.M.K., D.T.S.).

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare competing financial interests: details accompany the full-text HTML version of the paper at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to D.T.S. ([dscadden@mgh.harvard.edu](mailto:dscadden@mgh.harvard.edu)).



## METHODS

**Flow cytometric, CFU-C and LTC-IC analyses.** These analyses were as described previously<sup>19,20,21</sup>.

**Induction of deletion.** *Mx1* promoter was induced by three intra-peritoneal injections of polyI:C (Amersham Biosciences) 250 µg over 5 days. Deletion was confirmed by PCR.

**Engraftment studies.** For the competitive transplantation experiments,  $2.5 \times 10^5$  bone-marrow MNCs were mixed with  $2.5 \times 10^5$  bone-marrow MNC cells from wild-type mice and relative contributions to haematopoiesis assessed as described previously<sup>21</sup>.

**In vitro transmigration.** Chemotaxis assays used 5 µm pore transwells (Corning-Costar Corp.). Bone marrow cells or lymph node lymphocytes ( $5 \times 10^4$ ) were added to the upper well. Chemotaxis towards murine SDF-1α (PeproTech Inc.) in the lower chamber was scored at 3 h.

**Homing in vivo.** Purified bone-marrow cells or lymph node lymphocytes were labelled with 5 µM DiI or DiO (Molecular Probes Inc.) in accordance with the manufacturer's instructions. Cells were injected by tail vein and assessed by flow cytometry of bone marrow and spleen at 6 h, and of lymph nodes at 24 h.

**Cholera toxin treatment.** Purified cells were resuspended in fully supplemented medium at  $1 \times 10^6$  cells ml<sup>-1</sup>. Cholera toxin (10 µg ml<sup>-1</sup>; Sigma) was added, incubated at 37 °C for 1 h, washed three times with PBS and used in the *in vivo* homing and engraftment studies.

**In vivo microscopy.** Two-photon microscopy was performed as previously described<sup>12</sup>. Video imaging was performed 30–40 min after intravenous injection of labelled cells for 10 min; supplementary movies represent a 10-s window in the middle of the recording period. Still images of cells were obtained 60 min after injection and cells were quantified as stably homed if no movement was evident over a several minute observation interval.

19. Calvi, L. M. *et al.* Osteoblastic cells regulate the haematopoietic stem cell niche. *Nature* **425**, 841–846 (2003).
20. Cheng, T. *et al.* Hematopoietic stem cell quiescence maintained by p21<sup>cip1/waf1</sup>. *Science* **287**, 1804–1808 (2000).
21. Stier, S., Cheng, T., Dombkowski, D., Carlesso, N. & Scadden, D. T. Notch1 activation increases hematopoietic stem cell self-renewal *in vivo* and favors lymphoid over myeloid lineage outcome. *Blood* **99**, 2369–2378 (2002).

## LETTERS

# Histone modifications at human enhancers reflect global cell-type-specific gene expression

Nathaniel D. Heintzman<sup>1,2\*</sup>, Gary C. Hon<sup>1,3\*</sup>, R. David Hawkins<sup>1\*</sup>, Pouya Kheradpour<sup>5</sup>, Alexander Stark<sup>5,6</sup>, Lindsey F. Harp<sup>1</sup>, Zhen Ye<sup>1</sup>, Leonard K. Lee<sup>1</sup>, Rhona K. Stuart<sup>1</sup>, Christina W. Ching<sup>1</sup>, Keith A. Ching<sup>1</sup>, Jessica E. Antosiewicz-Bourget<sup>7</sup>, Hui Liu<sup>8</sup>, Xinmin Zhang<sup>8</sup>, Roland D. Green<sup>8</sup>, Victor V. Lobanikov<sup>9</sup>, Ron Stewart<sup>7</sup>, James A. Thomson<sup>7,10</sup>, Gregory E. Crawford<sup>11</sup>, Manolis Kellis<sup>5,6</sup> & Bing Ren<sup>1,4</sup>

The human body is composed of diverse cell types with distinct functions. Although it is known that lineage specification depends on cell-specific gene expression, which in turn is driven by promoters, enhancers, insulators and other *cis*-regulatory DNA sequences for each gene<sup>1–3</sup>, the relative roles of these regulatory elements in this process are not clear. We have previously developed a chromatin-immunoprecipitation-based microarray method (ChIP-chip) to locate promoters, enhancers and insulators in the human genome<sup>4–6</sup>. Here we use the same approach to identify these elements in multiple cell types and investigate their roles in cell-type-specific gene expression. We observed that the chromatin state at promoters and CTCF-binding at insulators is largely invariant across diverse cell types. In contrast, enhancers are marked with highly cell-type-specific histone modification patterns, strongly correlate to cell-type-specific gene expression programs on a global scale, and are functionally active in a cell-type-specific manner. Our results define over 55,000 potential transcriptional enhancers in the human genome, significantly expanding the current catalogue of human enhancers and highlighting the role of these elements in cell-type-specific gene expression.

We performed ChIP-chip analysis as described previously<sup>5</sup> to determine binding of CTCF (insulator-binding protein) and the coactivator p300 (also known as EP300), and patterns of histone modifications in five human cell lines: cervical carcinoma HeLa, immortalized lymphoblast GM06690 (GM), leukaemia K562, embryonic stem cells (ES) and BMP4-induced ES cells (dES). We first investigated 1% of the human genome selected by the ENCODE consortium<sup>7</sup>, using DNA microarrays consisting of 385,000 50-base oligonucleotides that tile 30-million base pairs (bp) at 36 bp resolution. We examined mono- and tri-methylation of histone H3 lysine 4 (H3K4me1, H3K4me3) and acetylation of histone H3 lysine 27 (H3K27ac) at well-annotated promoters, reasoning that the state of these histone modifications would vary in a cell-type-specific manner. To our surprise, the chromatin signatures at promoters are remarkably similar across all cell types (Fig. 1a). Quantitative comparison of ChIP-chip enrichment (see Supplementary Information) revealed highly correlated histone modification patterns at promoters across all cell types, with an average Pearson correlation coefficient of 0.71 (Supplementary Fig. 1a). This observation also holds for the larger set of GenCode promoters (Supplementary Fig. 2).

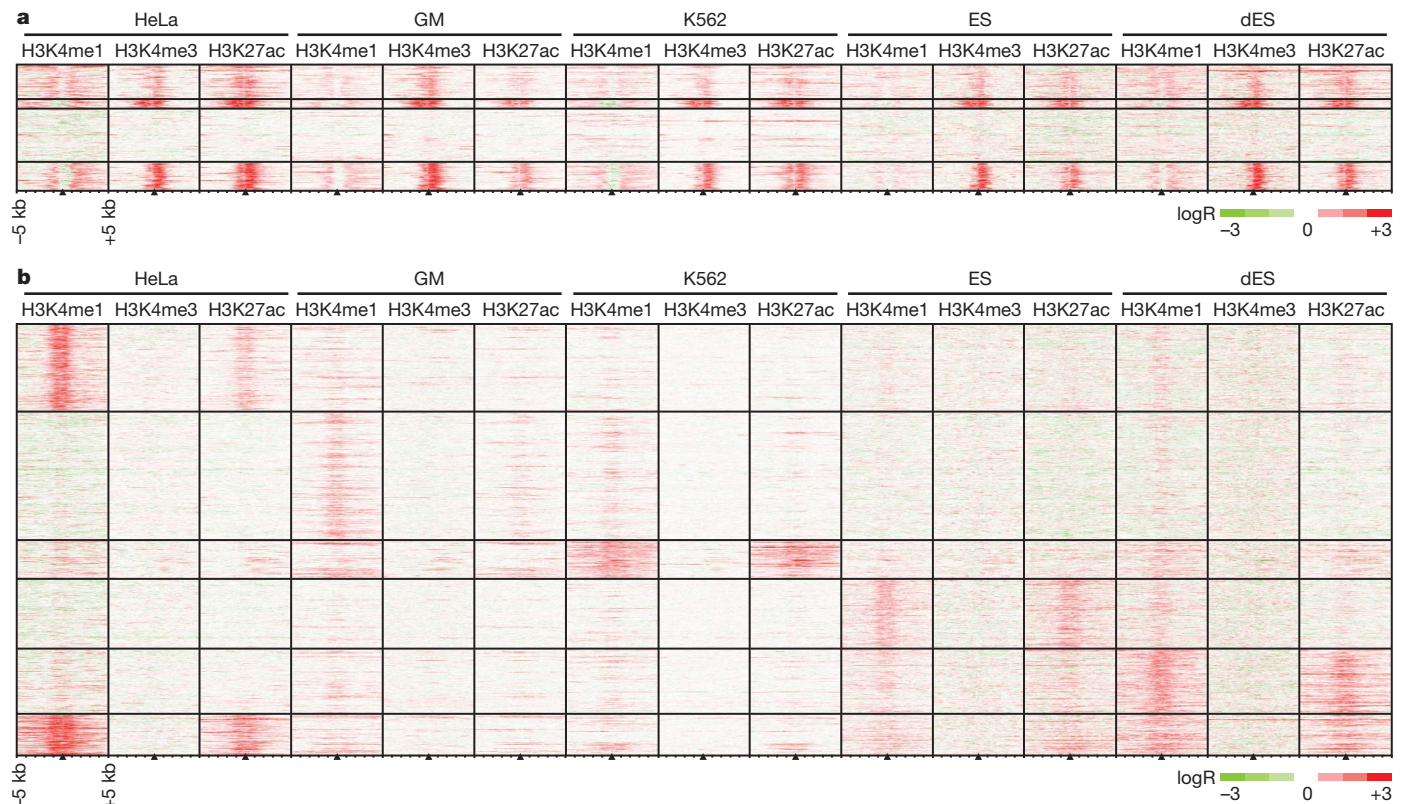
Next, we identified putative insulators in the ENCODE regions for these cell types based on CTCF binding, because mammalian insulators are generally understood to require CTCF to block promoter–enhancer interactions<sup>3</sup>. We observed nearly identical CTCF occupancy (Supplementary Table 1 and Supplementary Fig. 1e) and highly correlated CTCF enrichment patterns across all five cell types (Supplementary Fig. 1b), providing experimental support for the mostly cell-type-invariant function of CTCF as suggested by DNase hypersensitivity mapping results<sup>8</sup>.

We then investigated transcriptional enhancers in the ENCODE regions, performing ChIP-chip in HeLa, K562 and GM cells to locate binding sites for the transcriptional coactivator protein p300 (Supplementary Tables 2–4) because p300 is known to localize at enhancers<sup>9</sup>. We observed highly cell-type-specific histone modification patterns at distal p300-binding sites (Supplementary Fig. 1f), in marked contrast to the similarities in histone modifications across cell types at promoters. We then used our chromatin-signature-based prediction method<sup>5</sup> to identify additional enhancers in the ENCODE regions in these cell types (Fig. 1b and Supplementary Tables 5–9). In addition to the characteristic H3K4me1 enrichment, predicted enhancers are frequently marked by acetylation of H3K27, DNaseI hypersensitivity and/or binding of transcription factors and coactivators, and many contain evolutionarily conserved sequences (Supplementary Figs 3 and 4; see Supplementary Information). Unlike promoters and insulators, but similar to p300-binding sites, the histone modification patterns at predicted enhancers are largely cell-type-specific (Fig. 1b and Supplementary Fig. 1d), in agreement with observations that H3K4me1 is distributed in a cell-type-specific manner<sup>10</sup>.

These results indicate that enhancers are the most variable class of transcriptional regulatory element between cell types and are probably of primary importance in driving cell-type-specific patterns of gene expression. Knowledge of enhancers is therefore critical for understanding the mechanisms that control cell-type-specific gene expression, yet our incomplete knowledge of enhancers in the human genome has confined previous studies of gene regulatory networks mainly to promoters. To identify enhancers on a genome-wide scale and facilitate global analysis of gene regulatory mechanisms, we performed ChIP-chip throughout the entire human genome as described<sup>6</sup>, mapping enrichment patterns of H3K4me1 and H3K4me3 in HeLa cells. Using previously described chromatin signatures for enhancers<sup>5</sup>, we

<sup>1</sup>Ludwig Institute for Cancer Research, <sup>2</sup>Biomedical Sciences Graduate Program, <sup>3</sup>Bioinformatics Program, and <sup>4</sup>Department of Cellular and Molecular Medicine, UCSD School of Medicine, 9500 Gilman Drive, La Jolla, California 92093-0653, USA. <sup>5</sup>MIT Computer Science and Artificial Intelligence Laboratory, 32 Vassar Street, Cambridge, Massachusetts 02139, USA. <sup>6</sup>Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, Massachusetts 02142, USA. <sup>7</sup>Morgridge Institute for Research, Madison, Wisconsin 53707-7365, USA. <sup>8</sup>Roche NimbleGen, Inc., 500 South Rosa Road, Madison, Wisconsin 53719, USA. <sup>9</sup>National Institutes of Allergy and Infectious Disease, 5640 Fishers Lane, Rockville, Maryland 20852, USA. <sup>10</sup>University of Wisconsin School of Medicine and Public Health, Madison, Wisconsin 53706, USA. <sup>11</sup>Institute for Genome Sciences and Policy, and Department of Pediatrics, Duke University, 101 Science Drive, Durham, North Carolina 27708, USA.

\*These authors contributed equally to this work.



**Figure 1 | Chromatin modifications at promoters are generally cell-type-invariant whereas those at enhancers are cell-type-specific.** We used ChIP-chip to map histone modifications (H3K4me1, H3K4me3 and H3K27ac) in the ENCODE regions in five cell types (HeLa, GM, K562, ES, dES). **a**, We performed *k*-means clustering on the chromatin modifications found  $\pm 5$  kb from 414 promoters, and observe them to be generally

invariant across cell types. **b**, As in **a**, but clustering on 1,423 non-redundant enhancers predicted on the basis of chromatin signatures, revealing the cell-type-specificity of enhancers. LogR is the log ratio of enrichment of each marker as determined by ChIP-chip. Promoters and predicted enhancers are located at the centre of 10-kb windows as indicated by black triangles.

predicted 36,589 enhancers in the HeLa genome (Fig. 2a and Supplementary Table 10; see Supplementary Information). This method correctly located several previously characterized enhancers, including the  $\beta$ -globin HS2 enhancer<sup>11</sup> and distal enhancers for the PAX6 (ref. 12) and PLAT<sup>13</sup> genes (Fig. 2b). Most predicted enhancers are distal to promoters (Fig. 2c), have strong evolutionary conservation (see Supplementary Information) and are marked by histone acetylation (H3K27ac), binding of coactivator proteins (p300, MED1) or DNaseI hypersensitivity (DHS) (Fig. 2a, d; see Supplementary Information). We verified the functional potential of predicted HeLa enhancers using luciferase reporter assays as described<sup>5</sup> (see Supplementary Methods). Out of nine predicted enhancers that we evaluated, seven (78%) were active in reporter assays (Fig. 2e and Supplementary Table 11), with median activity significantly different from that in random genomic regions ( $P = 3.25 \times 10^{-4}$ ). These results support the suitability of using chromatin signatures to identify genomic regions with enhancer function.

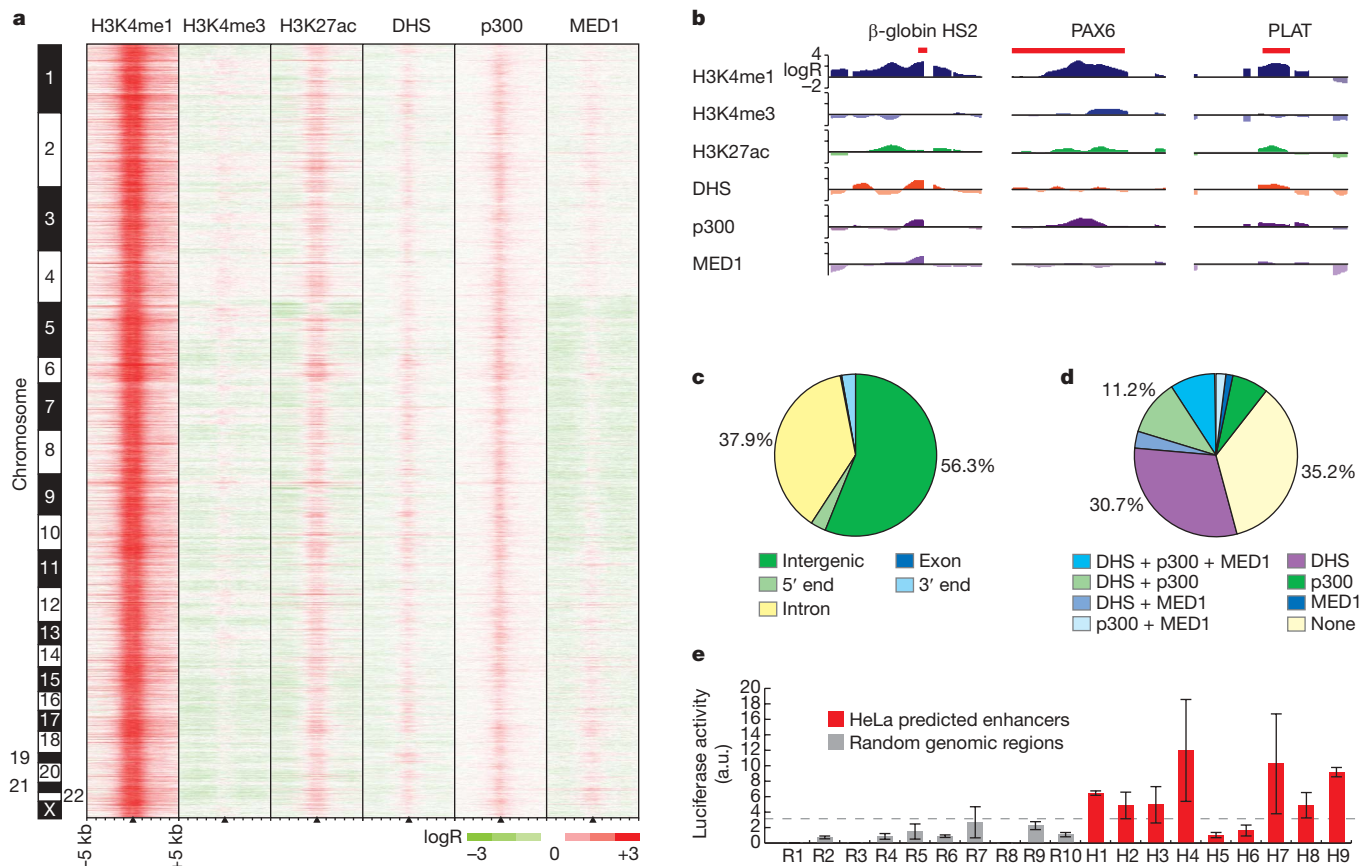
We evaluated the predicted enhancers for conserved motif-like sequences using several-hundred shuffled TRANSFAC motifs across ten mammals in a phylogenetic framework that tolerates motif movement, partial motif loss and sequencing or alignment discrepancies (see Supplementary Methods). Predicted enhancers showed conservation for 4.3% of instances (at branch-length-score  $> 50\%$ , see Supplementary Methods), which is substantially greater than for the remaining intergenic regions (2.9%,  $P < 1 \times 10^{-100}$ ) and even promoter regions (3.9%,  $P = 1 \times 10^{-57}$ ). Additionally, testing a list of 123 unique TRANSFAC motifs as described<sup>14</sup> (see Supplementary Information), we found that 67 (54%) are over-conserved and 39 (32%) are enriched in predicted enhancers (Supplementary Table 12). We also performed *de novo* motif discovery in enhancer regions using multiple alignments of 10 mammalian genomes (see Supplementary Methods),

revealing 41 enhancer motifs, 19 of which match known transcription factor motifs whereas 22 are new (Supplementary Table 13). These motifs show conservation rates between 7% and 22% in enhancers (median 9.3%), compared to only 1.1% for control shuffled motifs of identical composition. Furthermore, over 90% of these motifs seem to be unique to enhancers, as only 4 motifs are enriched in promoter regions and 12 are in fact depleted in promoters (Supplementary Table 13), indicating that predicted enhancers contain unique regulatory sequences that may be specific to enhancer function.

To investigate the association of predicted enhancers with HeLa-specific gene expression, we used Shannon entropy<sup>15</sup> to rank genes by the specificity of their expression levels in HeLa compared to that in three other cell lines (K562, GM06990, IMR90) (Supplementary Fig. 5; see Supplementary Information), and then plotted the distribution of enhancers around genes within insulator-delineated domains (as defined by CTCF-binding sites in Supplementary Fig. 6; see Supplementary Information). Predicted enhancers are markedly enriched near HeLa-specific expressed genes (Fig. 3a), particularly within 200 kb of promoters. We observed a 1.83-fold enrichment ( $P = 4.71 \times 10^{-279}$ ) of predicted enhancers around HeLa-specific expressed genes relative to a random distribution (see Supplementary Information) and significant depletion of enhancers around non-specific expressed genes ( $P = 5.43 \times 10^{-15}$ ) and HeLa-specific repressed genes ( $P = 4.63 \times 10^{-2}$ ).

To investigate more directly the relationship between chromatin modification patterns at enhancers and cell-type-specific gene expression, we expanded our global analysis to another cell type. We performed genome-wide ChIP-chip for H3K4me1 and H3K4me3 in K562 cells and identified 24,566 putative enhancers in this cell type using our chromatin-signature-based enhancer-prediction method (Supplementary Table 14; see Supplementary Information).





**Figure 2 | Genome-wide enhancer predictions in human cells.** **a**, We predict 36,589 enhancers in HeLa cells on the basis of chromatin signatures for H3K4me1 and H3K4me3 as determined by ChIP-chip using genome-wide tiling microarrays and condensed enhancer microarrays (see Supplementary Information). Enhancer predictions are located at the centre of 10-kb windows as indicated by black triangles, and ordered by genomic position. Enrichment data are shown for histone modifications (H3K4me1, H3K4me3 and H3K27ac), DNaseI hypersensitivity (DHS), and binding of p300 and MED1. **b**, ChIP-chip enrichment profiles at several known enhancers (indicated in red) recovered by prediction:  $\beta$ -globin HS2 (chromosome 11: 5,258,371–5,258,665)<sup>11</sup>, PAX6 (chromosome 11: 31,630,500–31,635,000)<sup>12</sup>, PLAT (chromosome 8: 42,191,500–42,192,400)<sup>13</sup> (5-kb windows centred on enhancer predictions; images generated in part at the UCSC Genome Browser). **c**, Predicted enhancer distribution relative to UCSC Known

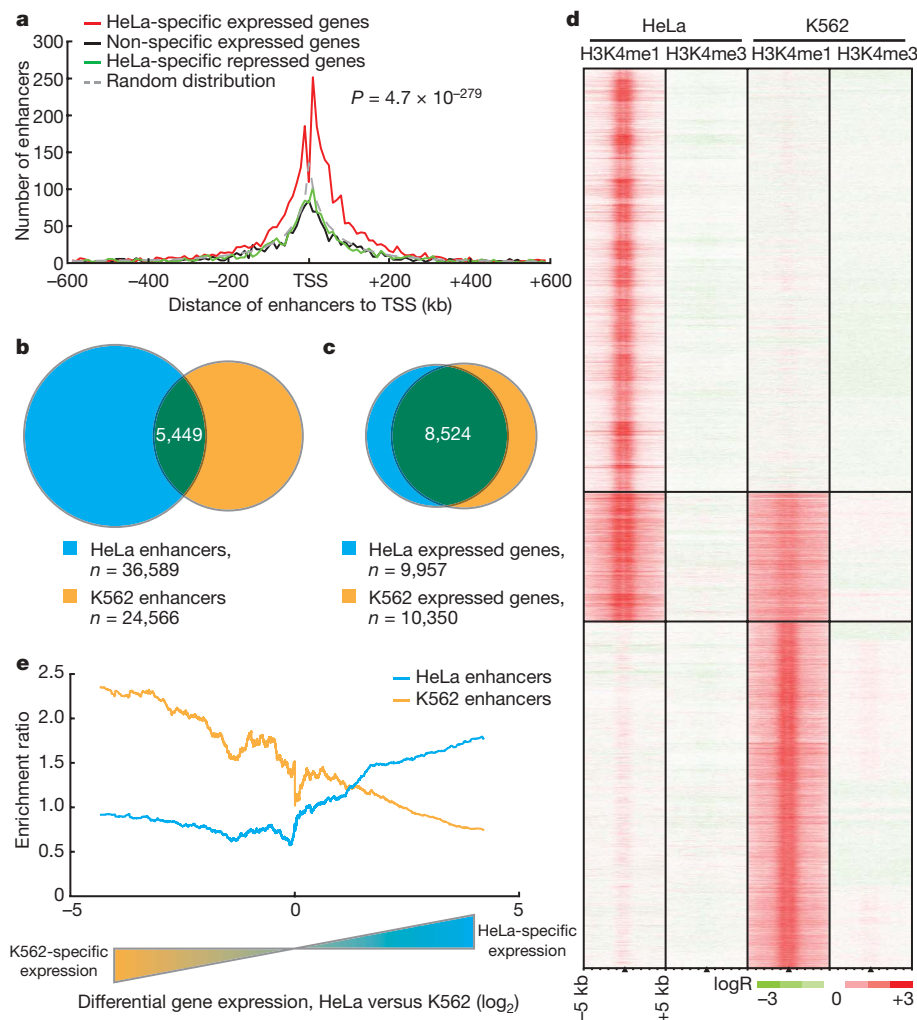
Consistent with results in the ENCODE regions, most enhancers predicted in K562 and HeLa cells are unique to either cell type (Fig. 3b), even though most expressed genes are common between the cell types (Fig. 3c). Chromatin modification profiles at predicted enhancers throughout the genome are highly cell-type-specific (Fig. 3d), with a Pearson correlation coefficient of  $-0.32$ . Furthermore, these differences seem to have regulatory implications, because domains with HeLa-specific expressed genes are enriched in HeLa enhancers but depleted in K562 enhancers, and vice versa (Fig. 3e; see Supplementary Information). These observations hold across all five cell types in the ENCODE regions (see Supplementary Information). To assess the cell-type-specificity of enhancer activity, we cloned enhancers predicted specifically in K562 cells (and not in HeLa cells) and subjected them to reporter assays in HeLa cells as described above. Out of nine K562-specific enhancers tested, only two (22%) were active in HeLa cells (Supplementary Fig. 7), and the median activity of the K-562-specific enhancers was not significantly different from random ( $P = 0.11$ ), indicating that the enhancer chromatin signature is a reliable marker of cell-type-specific enhancer function.

Although most enhancers are cell-type-specific, the presence of predicted enhancers shared by HeLa and K562 (Fig. 3b, d) indicates

Genes. Most enhancers have intergenic (56.3%) or intronic (37.9%) localization relative to UCSC Known Gene 5'-ends. **d**, Most enhancers (64.8%) are significantly marked by DNaseI hypersensitivity, binding of p300, binding of MED1, or some combination thereof. **e**, Seven out of nine enhancers predicted in HeLa cells were active in reporter assays (red bars) as compared to none of the random fragments selected as controls (grey), where activity is defined as relative luciferase value greater than 2.33 standard deviations ( $P = 0.01$ ) above the median random activity (grey dashed line). Error bars, standard deviation. Regions of  $\sim 1$ –2 kb in size were randomly selected for validation in reporter assays based on histone modification patterns as in **a**, overlap with features in **d**, and sequence features amenable to cloning by means of polymerase chain reaction (see Supplementary Information). a.u., arbitrary units.

that some enhancers may be active in multiple cell types or conditions. We compared the HeLa enhancer predictions with the results of several genome-wide studies of binding sites for sequence-specific transcription factors in different cell types, namely oestrogen receptor<sup>16</sup> (ER, also known as ESR1), p53 (TP53; ref. 17) and p63 (TP63; ref. 18) in MCF7, HCT116 and ME180 cells, respectively. Interestingly, significant percentages of binding sites for each transcription factor (from 21.4% to 32.6%) overlap with predicted enhancers in HeLa cells (Fig. 4a and Supplementary Table 15), in contrast to a significant depletion of the repressor NRSF (also known as REST)<sup>19</sup> at predicted enhancers and minimal overlap with CTCF-binding sites (see Supplementary Information).

To examine the potential role of enhancers in regulating inducible gene expression, we treated HeLa cells with the cytokine interferon- $\gamma$  (IFN- $\gamma$ ) and identified binding sites for the transcription factor STAT1 throughout the genome using ChIP-chip. STAT1 generally binds its target DNA sequences only after IFN- $\gamma$  induction<sup>20</sup>, with a small fraction of binding possible before induction<sup>21</sup>. In IFN- $\gamma$ -treated HeLa cells, we identified 1,969 STAT1-binding sites (Supplementary Table 16), with 85.8% of STAT1-binding sites occurring distal to Known Gene 5'-ends. Comparison of these distal STAT1-binding sites with recent

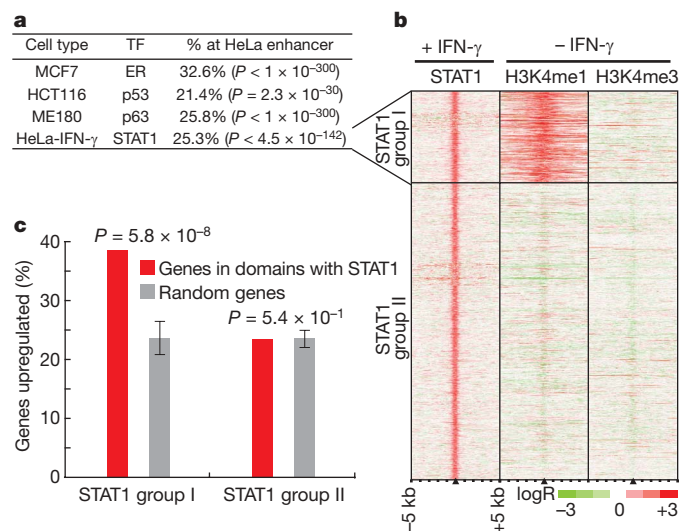


**Figure 3 | Chromatin modifications at enhancers are globally related to cell-type-specific gene expression.** **a**, Enhancer localization relative to HeLa-specific expressed genes compared to K562, GM06990 and IMR90 cells (red), non-specific expressed (green), HeLa-specific repressed (black), and a random distribution (dashed grey). Predicted enhancers are enriched around HeLa-specific expressed genes within insulator-defined domains and depleted in domains of ubiquitous or non-expressed genes ( $P$ -value reflects significance of enhancer enrichment in domains of HeLa-specific expressed genes, see Supplementary Information). TSS, transcription start site. **b, c**, Most enhancers predicted in HeLa and K562 cells are cell-type-specific (**b**) whereas most genes in HeLa and K562 cells are not specifically expressed (**c**);  $n$  = integer number of enhancers or genes in each set. **d**, Chromatin modification patterns are cell-type-specific at most of the 55,454 enhancers predicted in HeLa and K562 cells. **e**, Comparison of enhancer enrichment and differential gene expression between HeLa cells and K562 cells revealed that HeLa enhancers are enriched near HeLa-specific expressed genes (blue line) whereas K562 enhancers are enriched near K562-specific expressed genes (orange line).

analysis of STAT1 binding in uninduced HeLa cells<sup>21</sup> shows that only 6.5% of IFN- $\gamma$ -induced STAT1-binding sites are occupied by STAT1 before induction. We observed that 429 distal STAT1-binding sites overlapped enhancers predicted in HeLa cells before induction (Fig. 4a and Supplementary Table 15). The H3K4me1 enhancer chromatin signature is present before induction at these STAT1-binding sites, which we designated as STAT1 group I, whereas no evidence of this signature was visible at the remaining 1,260 distal STAT1-binding sites, designated STAT1 group II (Fig. 4b). Intriguingly, we

observed significant relative induction of expression of genes in the domains of STAT1-group-I-binding sites after just 30 min of IFN- $\gamma$  induction, whereas induction levels remained relatively unchanged for genes in the domains of other distal STAT1-group-II-binding sites during this time (Fig. 4c). These findings indicate that an enhancer chromatin signature confers increased regulatory responsiveness to a STAT1-binding site, in agreement with our previous discovery of functional enhancers in HeLa cells that were marked by the enhancer chromatin signature but were not active until they were bound by STAT1 (ref. 5).

Our findings offer, to our knowledge, the first genome-wide evaluation of the relationship between chromatin modifications at transcriptional enhancers and global programs of cell-type-specific gene expression. We determined over 55,000 potential enhancers in the human genome and show that the chromatin modifications at the



**Figure 4 | Chromatin modifications are associated with an increased regulatory response of transcription-factor-binding sites at enhancers.** **a**, Predicted enhancers in steady-state HeLa cells overlap with significant fractions of transcription-factor-binding sites (ER, p53, p63) in diverse cell types (MCF7, HCT116, ME180), as well as with STAT1-binding sites in HeLa cells treated with the cytokine interferon- $\gamma$  (HeLa-IFN- $\gamma$ ) (TF, transcription factor; TFBS, transcription factor binding sites). **b**, Hundreds of STAT1-binding sites after treatment (+ IFN- $\gamma$ ) are marked by the enhancer chromatin signature in HeLa cells even before treatment (- IFN- $\gamma$ ). **c**, In HeLa cells treated with IFN- $\gamma$  (upper panel), gene expression is significantly ( $P = 5.8 \times 10^{-8}$ ) more likely to be induced by STAT1 binding at sites with the enhancer chromatin signature (red, STAT1 group I) than by STAT1 binding at other distal sites (red, STAT1 group II) relative to a random distribution (grey). Error bars, standard deviation.

enhancers correlate with cell-type-specific gene expression and functional enhancer activity. Perhaps the most intriguing observation is the large number of enhancers identified from the investigation of just two cell lines. Because enhancers are mostly cell-type-specific, our data indicate the existence of a vast number of enhancers in the human genome, on the order of  $10^5$ – $10^6$ , that are used to drive specific gene expression programs in the 200 cell types of the human body. Future experiments with diverse cell types and experimental conditions will be necessary to comprehensively identify these regulatory elements and understand their roles in the specific gene expression program of each cell type.

## METHODS SUMMARY

HeLa, K562 and IMR90 cells were obtained from ATCC. GM06990 cells were acquired from Coriell. All cells were cultured under recommended conditions. Passage 32 H1 cells were cultured as described<sup>22</sup> with/without 200 ng ml<sup>-1</sup> BMP4 for 6 days (RND Systems). Chromatin preparation, ChIP, DNA purification and ligation-mediated PCR were performed as described using commercially available and custom-made antibodies, and ChIP samples were hybridized to tiling microarrays and to custom-made condensed enhancer microarrays (NimbleGen Systems, Inc.) as described<sup>5,6</sup>. DNase-chip was performed and the data analysed as described<sup>23</sup>. Cloning and reporter assays were performed as described<sup>5</sup>. Data were normalized as described<sup>5</sup>, and ChIP-chip targets for CTCF, p300, MED1 and STAT1 were selected with the Mpeak program. We used MA2C (ref. 24) to normalize and call peaks on Nimblegen HD2 arrays. Enhancers were predicted, and *k*-means clustering, intersection analysis and evolutionary conservation analysis were performed as described<sup>5</sup>. Motif analysis was performed as described<sup>25</sup>. Gene expression was analysed using HGU133 Plus 2.0 microarrays (Affymetrix) as described<sup>5</sup>. Specificity of expression was determined using a function of Shannon entropy<sup>15</sup>. We use the MAS5 algorithm from the Bioconductor R package to generate gene expression present/absent calls. Detailed methods can be found in Supplementary Information. Supplementary data for the microarray experiments have been formatted for viewing in the UCSC genome browser via <http://bioinformatics-renlab.ucsd.edu/enhancer>.

Received 17 October 2008; accepted 26 January 2009.

Published online 18 March 2009.

1. Heintzman, N. D. & Ren, B. The gateway to transcription: identifying, characterizing and understanding promoters in the eukaryotic genome. *Cell. Mol. Life Sci.* **64**, 386–400 (2007).
2. Nightingale, K. P., O'Neill, L. P. & Turner, B. M. Histone modifications: signalling receptors and potential elements of a heritable epigenetic code. *Curr. Opin. Genet. Dev.* **16**, 125–136 (2006).
3. Maston, G. A., Evans, S. K. & Green, M. R. Transcriptional regulatory elements in the human genome. *Annu. Rev. Genomics Hum. Genet.* **7**, 29–59 (2006).
4. Kim, T. H. *et al.* A high-resolution map of active promoters in the human genome. *Nature* **436**, 876–880 (2005).
5. Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature Genet.* **39**, 311–318 (2007).
6. Kim, T. H. *et al.* Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* **128**, 1231–1245 (2007).
7. ENCODE Consortium. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* **306**, 636–640 (2004).
8. Xi, H. *et al.* Identification and characterization of cell type-specific and ubiquitous chromatin regulatory structures in the human genome. *PLoS Genet.* **3**, e136 (2007).

9. Wang, Q., Carroll, J. S. & Brown, M. Spatial and temporal recruitment of androgen receptor and its coactivators involves chromosomal looping and polymerase tracking. *Mol. Cell* **19**, 631–642 (2005).
10. Koch, C. M. *et al.* The landscape of histone modifications across 1% of the human genome in five human cell lines. *Genome Res.* **17**, 691–707 (2007).
11. King, D. C. *et al.* Evaluation of regulatory potential and conservation scores for detecting *cis*-regulatory modules in aligned mammalian genome sequences. *Genome Res.* **15**, 1051–1060 (2005).
12. Kleinjan, D. A. *et al.* Aniridia-associated translocations, DNase hypersensitivity, sequence comparison and transgenic analysis redefine the functional domain of PAX6. *Hum. Mol. Genet.* **10**, 2049–2059 (2001).
13. Wolf, A. T., Medcalf, R. L. & Jern, C. The t-PA -7351C>T enhancer polymorphism decreases Sp1 and Sp3 protein binding affinity and transcriptional responsiveness to retinoic acid. *Blood* **105**, 1060–1067 (2005).
14. Xie, X. *et al.* Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. *Nature* **434**, 338–345 (2005).
15. Schug, J. *et al.* Promoter features related to tissue specificity as measured by Shannon entropy. *Genome Biol.* **6**, R33 (2005).
16. Carroll, J. S. *et al.* Genome-wide analysis of estrogen receptor binding sites. *Nature Genet.* **38**, 1289–1297 (2006).
17. Wei, C. L. *et al.* A global map of p53 transcription-factor binding sites in the human genome. *Cell* **124**, 207–219 (2006).
18. Yang, A. *et al.* Relationships between p63 binding, DNA sequence, transcription activity, and biological function in human cells. *Mol. Cell* **24**, 593–602 (2006).
19. Johnson, D. S., Mortazavi, A., Myers, R. M. & Wold, B. Genome-wide mapping of *in vivo* protein-DNA interactions. *Science* **316**, 1497–1502 (2007).
20. Brivanlou, A. H. & Darnell, J. E. Jr. Signal transduction and the control of gene expression. *Science* **295**, 813–818 (2002).
21. Robertson, G. *et al.* Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nature Methods* **4**, 651–657 (2007).
22. Ludwig, T. E. *et al.* Feeder-independent culture of human embryonic stem cells. *Nature Methods* **3**, 637–646 (2006).
23. Crawford, G. E. *et al.* DNase-chip: a high-resolution method to identify DNase I hypersensitive sites using tiled microarrays. *Nature Methods* **3**, 503–509 (2006).
24. Song, J. S. *et al.* Model-based analysis of two-color arrays (MA2C). *Genome Biol.* **8**, R178 (2007).
25. Kheradpour, P., Stark, A., Roy, S. & Kellis, M. Reliable prediction of regulator targets using 12 *Drosophila* genomes. *Genome Res.* **17**, 1919–1931 (2007).

Supplementary Information is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank members of the Ren laboratory for comments. This work was supported by funding from American Cancer Society (R.D.H.), NIAID Intramural Research Program (V.V.L.), LICR (B.R.), NHGRI (B.R.), NCI (B.R.) and CIRM (B.R.).

**Author Contributions** R.D.H., N.D.H., G.C.H. and B.R. designed the experiments; R.D.H., N.D.H., L.F.H., Z.Y., L.K.L., R.K.S., C.W.C., H.L. and X.Z. conducted the ChIP-chip experiments; G.C.H. and K.A.C. analysed the ChIP-chip data; G.C.H. predicted enhancers; R.D.H. and L.K.L. conducted the reporter assays; J.E.A.-B., R.S. and J.A.T. provided human ES cells and expression data; V.V.L. provided advice and antibodies for CTCF-ChIP experiments; P.K., A.S. and M.K. analysed the transcription factor motifs; G.E.C. performed and analysed the DNaseI-chip experiments; and N.D.H., G.C.H., R.D.H. and B.R. wrote the manuscript.

**Author Information** Microarray data have been submitted to the GEO repository under accession numbers GSE14083, GSE8098, GSE7872 and GSE7118. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to B.R. ([biren@ucsd.edu](mailto:biren@ucsd.edu)).



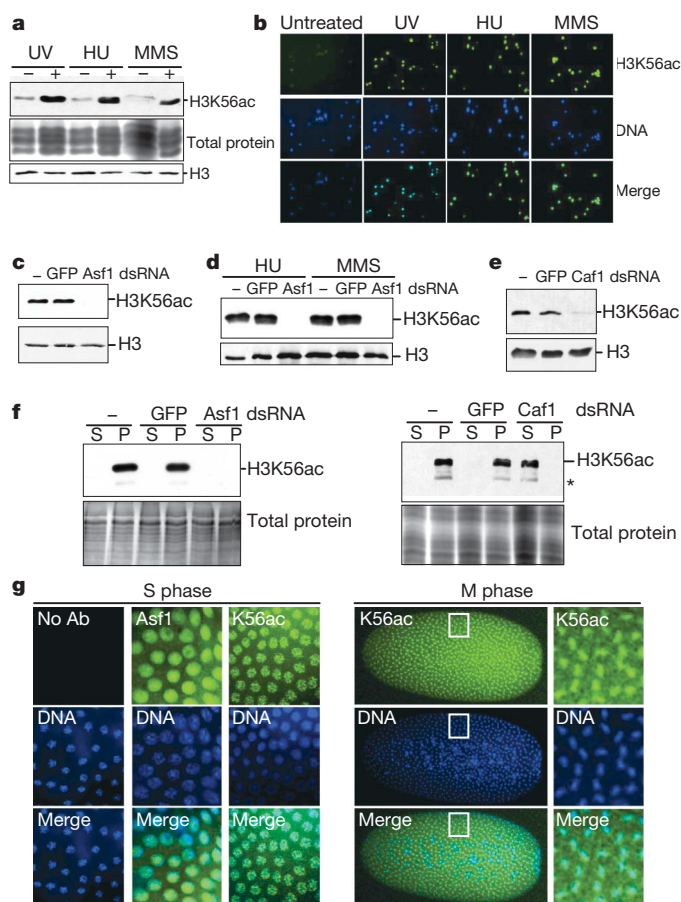
# CBP/p300-mediated acetylation of histone H3 on lysine 56

Chandrima Das<sup>1</sup>, M. Scott Lucia<sup>2</sup>, Kirk C. Hansen<sup>3</sup> & Jessica K. Tyler<sup>1</sup>

Acetylation within the globular core domain of histone H3 on lysine 56 (H3K56) has recently been shown to have a critical role in packaging DNA into chromatin following DNA replication and repair in budding yeast<sup>1,2</sup>. However, the function or occurrence of this specific histone mark has not been studied in multicellular eukaryotes, mainly because the Rtt109 enzyme that is known to mediate acetylation of H3K56 (H3K56ac) is fungal-specific<sup>3,4</sup>. Here we demonstrate that the histone acetyl transferase CBP (also known as Nejire) in flies and CBP and p300 (Ep300) in humans acetylate H3K56, whereas *Drosophila* Sir2 and human SIRT1 and SIRT2 deacetylate H3K56ac. The histone chaperones ASF1A in humans and Asf1 in *Drosophila* are required for acetylation of H3K56 *in vivo*, whereas the histone chaperone CAF-1 (chromatin assembly factor 1) in humans and Caf1 in *Drosophila* are required for the incorporation of histones bearing this mark into chromatin. We show that, in response to DNA damage, histones bearing acetylated K56 are assembled into chromatin in *Drosophila* and human cells, forming foci that colocalize with sites of DNA repair. Furthermore, acetylation of H3K56 is increased in multiple types of cancer, correlating with increased levels of ASF1A in these tumours. Our identification of multiple proteins regulating the levels of H3K56 acetylation in metazoans will allow future studies of this critical and unique histone modification that couples chromatin assembly to DNA synthesis, cell proliferation and cancer.

H3K56 acetylation has a critical role in regulating chromatin assembly following DNA synthesis<sup>1,2</sup>, chromatin disassembly during transcriptional activation<sup>5</sup> and cell survival<sup>1</sup> in yeast. Although H3K56ac clearly exists in *Drosophila*<sup>6,7</sup>, it is not known whether these functions of H3K56ac extend to multicellular eukaryotes. To investigate the function of H3K56ac in higher eukaryotes (Supplementary Fig. 1), we asked whether histones carrying K56ac are incorporated into chromatin following DNA repair in *Drosophila* S2 cells using antibodies specific to H3K56ac (Supplementary Fig. 2). Exposure to hydroxyurea, methyl methane sulphonate (MMS) or ultraviolet irradiation increased the level of H3K56ac on chromatin in a dose-dependent manner, as determined by western blotting (Fig. 1a and Supplementary Fig. 3) and immunofluorescence analyses (Fig. 1b and Supplementary Fig. 4). Notably, these agents did not result in an accumulation of cells in S phase (Supplementary Fig. 5). Consistent with the requirement for the histone chaperone anti-silencing function 1 (Asf1) for H3K56 acetylation in yeast<sup>8</sup>, we found that *Drosophila* Asf1 is required for both the endogenous level (Fig. 1c and Supplementary Fig. 6) and the DNA-damage-induced increase in the level (Fig. 1d) of H3K56ac *in vivo*. Notably, knockdown of another histone chaperone, chromatin assembly factor 1 (Caf1), also markedly decreased the levels of H3K56ac on chromatin (Fig. 1e). However, knockdown of Caf1, but not Asf1, resulted in accumulation of H3K56ac in the soluble protein fraction (Fig. 1f). As such,

these results demonstrate that *Drosophila* Caf1 is required for the assembly of histones carrying the K56ac mark into chromatin, whereas Asf1 is required for H3K56 acetylation *per se*. Using *Drosophila* embryos cycling synchronously between S and M phases,

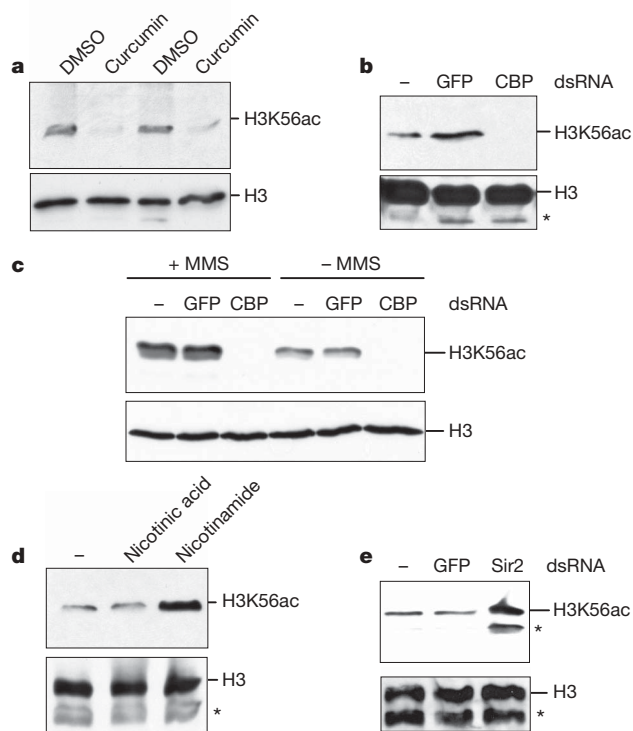


**Figure 1 | *Drosophila* Asf1 promotes H3K56 acetylation whereas Caf1 deposits H3K56ac into chromatin.** **a, b,** H3K56ac levels in chromatin after DNA damage. HU, hydroxyurea; UV, ultraviolet. **c,** Asf1 is required for H3K56 acetylation, and in response to DNA damage. 'GFP' indicates a control RNAi against the green fluorescent protein. **d, e,** H3K56ac levels on chromatin following Caf1 p180 knockdown. **f,** H3K56ac levels in free histones in the supernatant (S) and chromatin-bound pellet (P). The asterisk indicates a likely proteolytic product of H3. **g,** *Drosophila* embryos were stained with no primary antibody (No Ab) or with antisera to Asf1 or K56ac, as indicated. Secondary antibody was included in all analyses. Magnification of the boxed regions is shown to the right. Original magnification,  $\times 600$  (**b**) and  $\times 100$  (**g**).

<sup>1</sup>Department of Biochemistry and Molecular Genetics, <sup>2</sup>Department of Pathology, and <sup>3</sup>Cancer Center Proteomics Core, University of Colorado School of Medicine, PO Box 6511, Aurora Colorado 80045, USA.

we detected no apparent difference in the amount of H3K56ac present in S- or M-phase nuclei (Fig. 1g). During mitosis, the staining of H3K56ac closely followed that of the condensed mitotic chromosomes, indicating that the H3K56ac mark is indeed on the chromatin rather than free in the nucleus. This is in contrast to the situation in yeast, where high levels of H3K56ac are only detectable in S phase<sup>9</sup>.

The enzymes that acetylate and deacetylate H3K56 are unknown in multicellular organisms. Given the structural similarity between the yeast H3K56 histone acetyl transferase (HAT) Rtt109 (ref. 10) and p300/CBP<sup>11</sup>, we investigated the potential role of *Drosophila* CBP<sup>12</sup> in acetylating H3K56. Treatment of S2 cells with curcumin—an inhibitor of the CBP/p300 family of HAT proteins<sup>13</sup>—markedly decreased H3K56ac levels (Fig. 2a). Furthermore, knockdown of CBP (Fig. 2b and Supplementary Fig. 6), but not knockdown of another H3-specific HAT, Gcn5 (also known as Pcaf; ref. 14; Supplementary Fig. 6), blocked acetylation of H3K56 in both the absence and the presence of DNA damage (Fig. 2b, c), indicating that CBP is the major H3K56 acetylase in *Drosophila*. Consistent with the idea that Asf1 and CBP function together to acetylate H3K56 in flies, CBP coimmunoprecipitates with Asf1 (Supplementary Fig. 7). The NAD-dependent histone deacetylases (HDACs) Hst3 and Hst4 are the major H3K56ac HDACs in yeast<sup>15,16</sup>. To investigate whether an NAD-dependent HDAC mediates H3K56 deacetylation in *Drosophila*, we treated S2 cells with the NAD-dependent HDAC inhibitor nicotinamide<sup>17</sup>. Treatment with nicotinamide, but not with the related molecule nicotinic acid, resulted in increased H3K56ac levels (Fig. 2d and Supplementary Fig. 8). Of the six *Drosophila* NAD-dependent HDACs, Sir2 is the closest counterpart of yeast Hst3 and Hst4. Indeed, knockdown of *Drosophila* Sir2 greatly increased the level of H3K56ac (Fig. 2e and Supplementary Fig. 6), indicating that Sir2 deacetylates H3K56ac in *Drosophila*.

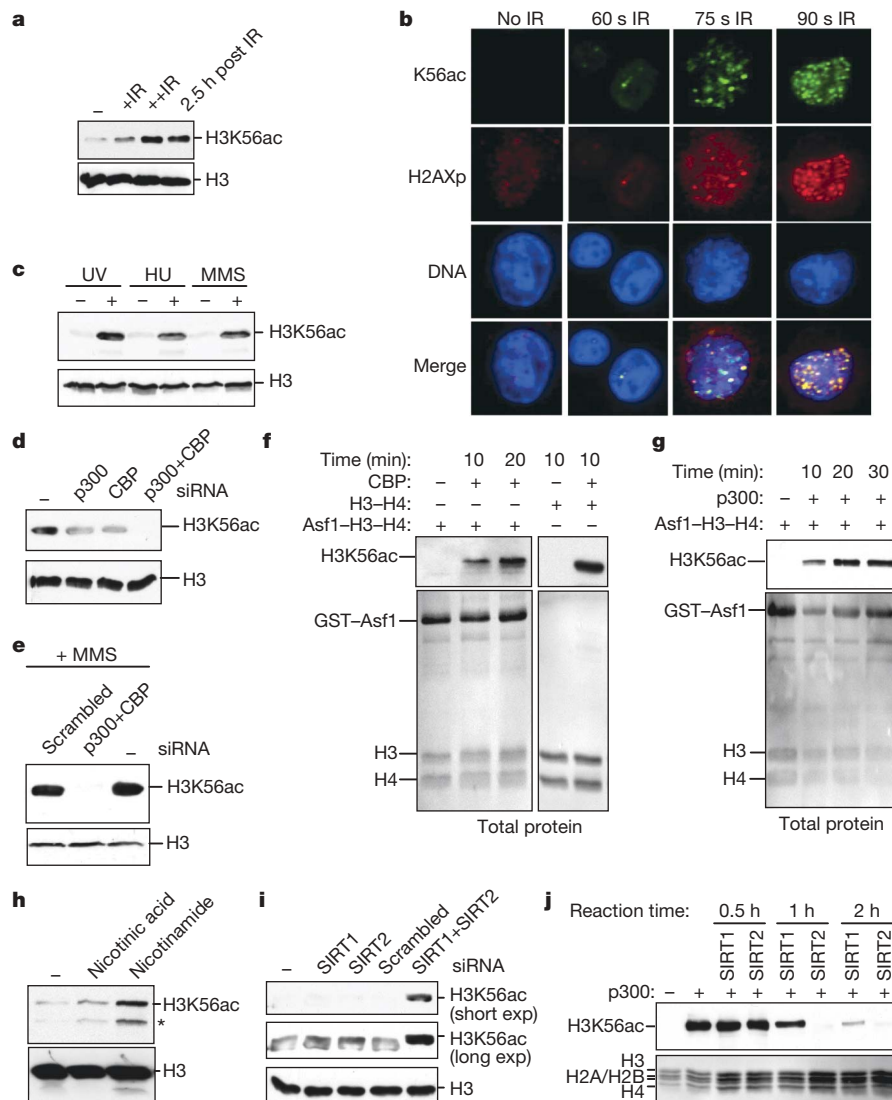


**Figure 2 | *Drosophila* CBP acetylates H3K56 whereas Sir2 deacetylates H3K56ac in vivo.** **a**, Inhibition of CBP by curcumin lowers levels of H3K56ac in S2 cells. Duplicate analyses are shown. **b**, CBP acetylates H3K56. The asterisk indicates a likely proteolytic product of H3. **c**, CBP is required for the increase of H3K56ac on chromatin after DNA damage. **d**, Inhibition of NAD-dependent HDACs by nicotinamide increases K56ac. **e**, Sir2 deacetylates H3K56ac

Besides one global mass spectrometry study of histone modifications<sup>18</sup>, H3K56ac has not previously been reported in human cells. By western blot analysis of chromatinized histones from HeLa cells, we clearly detected H3K56ac in humans (Fig. 3a). The amount of H3K56ac in HeLa and S2 cells was similar (Supplementary Fig. 9), and mass spectrometry analysis confirmed the existence of H3K56 acetylation in HeLa cells (Supplementary Fig. 10). Furthermore, the amount of H3K56ac on chromatin increased in response to gamma-irradiation in a dose-dependent manner (Fig. 3a). The clear colocalization of H3K56ac and the phosphorylated histone variant H2AX (H2AXp) following DNA damage indicated that the assembly of histones carrying H3K56ac is enriched at sites of DNA repair (Fig. 3b). The levels of H3K56ac on human chromatin were also increased following ultraviolet, MMS and hydroxyurea treatment (Fig. 3c). By flow cytometry analysis, we found that, even in the absence of DNA damage, human cells in all stages of the cell cycle have abundant H3K56ac staining (Supplementary Fig. 11).

Consistent with the reduced level of acetylation of H3K56 in human cells on curcumin treatment (Supplementary Fig. 12), we found that both CBP and p300 contribute to acetylation of H3K56 in human cells (Fig. 3d, e and Supplementary Fig. 13). Furthermore, CBP and p300 directly acetylate histone H3 on lysine 56 *in vitro* (Fig. 3f, g), with high efficiency as determined by mass spectrometry analysis (Supplementary Fig. 13). Acetylation of H3K56ac by the yeast HAT Rtt109 requires that histones H3–H4 be pre-bound to the histone chaperone Asf1 (ref. 19). However, this is not the case with CBP *in vitro*, because it efficiently acetylated H3K56 irrespective of whether the histones were bound to Asf1 (Fig. 3f). Towards identifying the human HDAC for H3K56, we found that treatment of human cells with nicotinamide greatly increased the level of H3K56ac (Fig. 3h), implicating an NAD-dependent HDAC. Small interfering RNA (siRNA) knockdown analyses of human SIRT1 and SIRT2 indicated that both enzymes contribute to deacetylation of H3K56 in human cells (Fig. 3i). Indeed, both recombinant human SIRT2 and SIRT1 deacetylated H3K56 *in vitro*, with SIRT2 being the more active enzyme (Fig. 3j and Supplementary Fig. 14). Treatment of human cells with sodium butyrate notably increased H3K56ac levels (Supplementary Fig. 12), additionally implicating a class I or class II HDAC in H3K56 deacetylation. Taken together, these data reveal that human CBP and p300 are the H3K56 acetylases whereas SIRT2 and SIRT1 are H3K56ac deacetylases.

Next, we investigated whether either of the two forms of the human ASF1 histone chaperone, ASF1A and ASF1B, is involved in H3K56 acetylation. Using yeast strains that express equivalent levels of human ASF1A, human ASF1B or *Drosophila* Asf1 in the place of the yeast Asf1 protein<sup>20</sup>, it was apparent that human ASF1A and *Drosophila* Asf1, but not human ASF1B, can assist yeast Rtt109 in the acetylation of H3K56 (Fig. 4a). In agreement, knockdown of ASF1A decreased the amount of H3K56ac in human cells more than knockdown of ASF1B whereas the combined knockdown of both ASF1A and ASF1B resulted in no detectable H3K56ac (Fig. 4b). The requirement for ASF1 for H3K56 acetylation *in vivo*, but not *in vitro*, indicates that additional constraints exist *in vivo*. We also found that human ASF1 is required for the increase in levels of H3K56ac on chromatin following DNA damage, whereas human CAF-1 is required for the assembly of the H3K56 acetylated histones onto DNA (Fig. 4c, d). Further supporting the role of ASF1A versus ASF1B in H3K56 acetylation, there is a notable correlation between levels of H3K56 acetylation and ASF1A but not ASF1B in a wide variety of normal and cancerous human tissues (Fig. 4e). The increase in levels of H3K56ac in many cancer tissues (Fig. 4e and Supplementary Figs 15 and 16) in a manner that is proportional to tumour grade (Supplementary Fig. 17 and Supplementary Table 1) led us to investigate whether H3K56ac levels correlate with proliferation and/or tumorigenicity. Non-tumorigenic MCF10A cells show much lower levels of H3K56ac than tumorigenic MCF7 breast cancer cells (Fig. 4f). Furthermore, H3K56ac staining also occurred in cancer



**Figure 3 | Human CBP and p300 acetylate H3K56 whereas SIRT1 and SIRT2 deacetylate H3K56.** **a**, Levels of H3K56ac on human chromatin following gamma irradiation (IR). + and ++ indicate increasing amounts of gamma irradiation. **b**, H3K56ac localizes to DNA damage foci. Original magnification,  $\times 600$ . **c**, Hydroxyurea (HU), MMS and ultraviolet (UV) induce H3K56ac. **d**, **e**, CBP and p300 are required for acetylation of H3K56 in the absence (**d**) and presence (**e**) of DNA damage. **f**, **g**, Human CBP (**f**) and

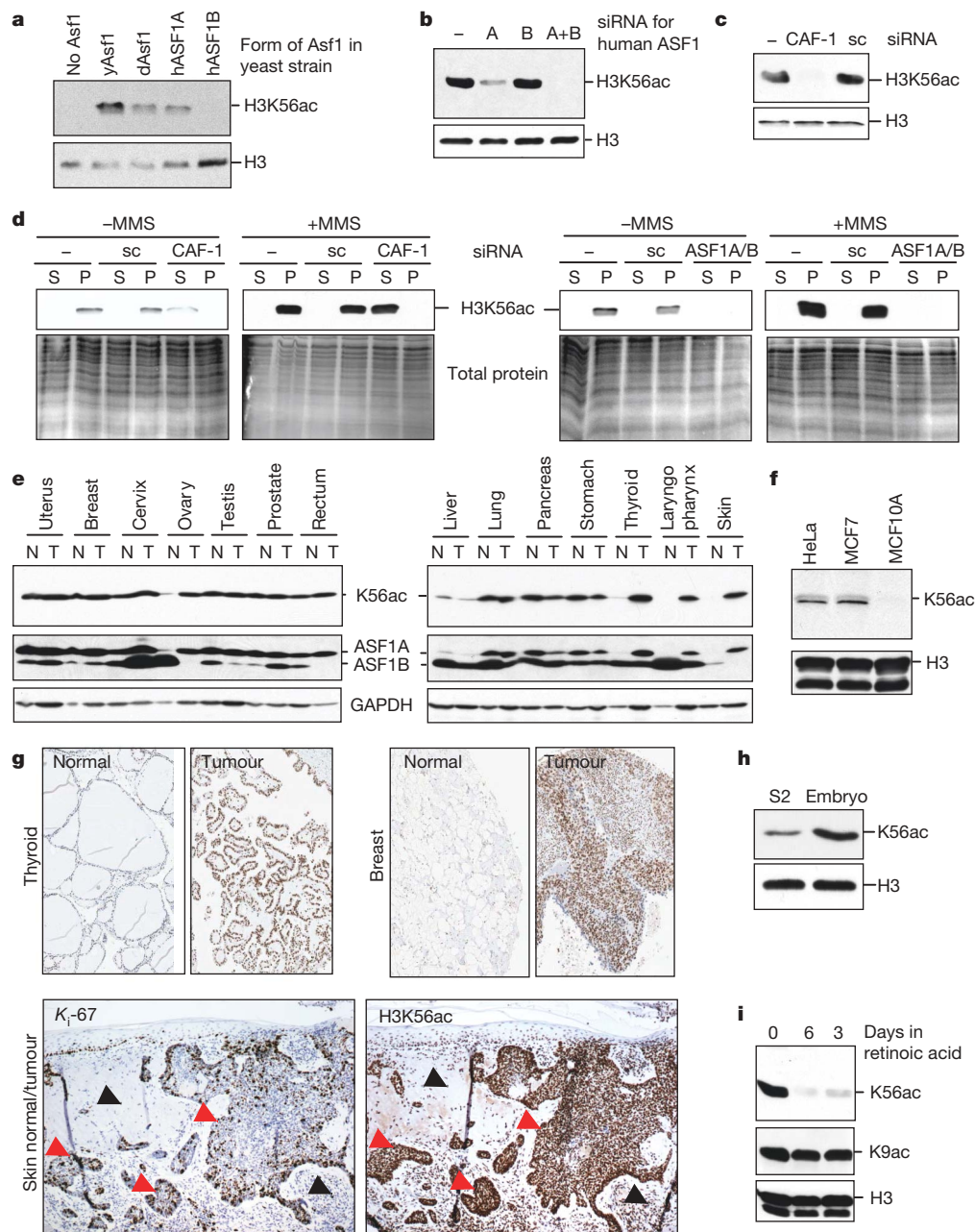
p300 (**g**) acetylate H3K56 *in vitro*. 'GST-Asf1' indicates recombinant Asf1 fused to the glutathione S-transferase affinity tag. **h**, Inhibition of NAD-dependent HDACs increases H3K56ac levels. The asterisk indicates a likely proteolytic product of H3. **i**, Human SIRT2 and SIRT1 deacetylate H3K56ac *in vivo*. Long and short exposures (exp) are shown. **j**, Human SIRT2 and SIRT1 deacetylate H3K56ac *in vitro*.

cells that are not positive for the proliferation marker  $K_i$ -67 (ref. 21; Fig. 4g). These data indicate that H3K56ac levels correlate with tumorigenicity rather than proliferation. Consistent with the increased level of H3K56ac in dedifferentiated cancer cells, we also find that H3K56ac levels positively correlate with the undifferentiated nature of cells. For example, H3K56ac levels are higher in embryos than in cell lines (Fig. 4h) and the level of H3K56ac rapidly drops on *in vitro* differentiation of cell lines (Fig. 4i and Supplementary Fig. 18).

Recent yeast studies indicate that H3K56ac increases the ability of histones to bind to CAF-1, which subsequently deposits the histones onto newly synthesized DNA<sup>1,2</sup>. This is also likely to be the case in humans, given that knockdown of CAF-1 blocks incorporation of H3K56ac into chromatin (Figs 1f and 4d) and that inhibition of ASF1 blocks DNA replication and chromatin assembly of newly synthesized DNA<sup>22–24</sup>. In agreement, CBP/p300 promote DNA replication and cell proliferation<sup>25,26</sup>. Furthermore, p300 is recruited to sites

of DNA synthesis via its interaction with PCNA<sup>27</sup>, consistent with a direct role for CBP/p300 in promoting chromatin assembly via K56 acetylation after DNA synthesis. It is also intriguing that the HAT activity of human CBP is highest at the G1/S boundary—the time at which massive amounts of newly synthesized H3 is produced to assemble the newly replicated genomes into chromatin<sup>28</sup>. We propose that the CBP/p300-mediated acetylation of H3K56 promotes the subsequent assembly of newly synthesized DNA into chromatin in metazoans (Supplementary Fig. 1), the inhibition of which negatively feeds back to block DNA replication and cell proliferation. Given that the H3K56ac mark is rapidly removed by histone deacetylases after histone incorporation onto newly replicated DNA in yeast<sup>15,16</sup>, it will be interesting to determine whether the H3K56ac deacetylation process is defective in undifferentiated and dedifferentiated cancer cells. Alternatively, the increased levels of H3K56ac in cancer cells and undifferentiated cells may reflect hyperdynamic exchange of histones<sup>29</sup> in undifferentiated cells that is lost on differentiation.





**Figure 4 | Asf1a drives K56 acetylation in human cells and tumours.**

**a**, ASF1A promotes K56 acetylation in yeast and humans (**b**). 'A' refers to ASF1A and 'B' refers to ASF1B. **c**, **d**, CAF-1 mediates H3K56ac assembly onto chromatin. sc, scrambled siRNA; P, chromatin-bound pellet; S, supernatant. **e**, K56ac and ASF1A correlate in matched normal (N) and tumour (T) pairs.

## METHODS SUMMARY

**Yeast strains and tissue culture.** Genotypes of strains and tissue culture details are described in the Methods.

**Histone isolation and western blotting.** Histone extraction was performed as described previously<sup>30</sup>. In brief, following nuclear pellet isolation, soluble chromatinized histones were extracted with 0.4 M and 0.25 M HCl for S2 cells and HeLa cells, respectively, followed by TCA precipitation. NP40-induced extraction of detergent-soluble proteins is detailed in the Methods. In brief, following treatment with 0.1% NP40 for 10 min, centrifugation at 1,800g, for 10 min led to the separation of supernatant and pellet fractions. Antibody details are provided in the Methods.

**Immunofluorescence.** Immunofluorescence following 0.5% Triton X-100 extraction was performed as described previously<sup>23</sup>.

**RNA interference.** Knockdowns were performed in S2 cells as described previously<sup>23</sup>. In brief, double-stranded RNA was generated by *in vitro* transcription using the Megascript kit from Ambion. After testing the quality of each double-stranded RNA, it was transfected into S2 cells for different lengths of time to ensure complete silencing

**f**, Increased K56ac in tumorigenic cells and tumours (**g**). Nuclei, blue; K56ac, brown. Comparison of K56ac and the proliferation marker Ki-67 in adjacent sections of late grade skin cancer. Red arrowheads, invading tumour; black arrowheads, normal tissue. Original magnification,  $\times 600$ . **h**, *Drosophila* embryos have increased K56Ac. **i**, Increased K56ac in undifferentiated cells.

of the gene of interest, as determined by western blotting. Primer sequences are available on request. siRNAs for silencing in HeLa cells were obtained from Dharmacon. **HAT and HDAC assays.** Recombinant human CBP, p300, SIRT1 and SIRT2 were obtained from Biomol. Following a 10 min preincubation of HAT with purified recombinant Asf1–H3H4 complexes, purified *Drosophila* core histones or H3–H4 tetramers at 30 °C, acetyl-CoA was added and incubated at 30 °C for 10 min, followed by TCA precipitation. For HDAC assays, following acetylation of H3K56 by CBP/p300, curcumin was added to inhibit CBP/p300 and the histones carrying H3K56ac were incubated with SIRT1 or SIRT2 in the presence of NAD at 30 °C for the indicated length of time.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 14 November 2008; accepted 2 February 2009.

Published online 8 March 2009.

- Chen, C. C. *et al.* Acetylated lysine 56 on histone H3 drives chromatin assembly after repair and signals for the completion of repair. *Cell* **134**, 231–243 (2008).

2. Li, Q. *et al.* Acetylation of histone H3 lysine 56 regulates replication-coupled nucleosome assembly. *Cell* **134**, 244–255 (2008).
3. Han, J. *et al.* Rtt109 acetylates histone H3 lysine 56 and functions in DNA replication. *Science* **315**, 653–655 (2007).
4. Xhemalce, B. *et al.* Regulation of histone H3 lysine 56 acetylation in *Schizosaccharomyces pombe*. *J. Biol. Chem.* **282**, 15040–15047 (2007).
5. Williams, S. K., Truong, D. & Tyler, J. K. Acetylation in the globular core of histone H3 on lysine-56 promotes chromatin disassembly during transcriptional activation. *Proc. Natl Acad. Sci. USA* **105**, 9000–9005 (2008).
6. Xu, F., Zhang, K. & Grunstein, M. Acetylation in histone H3 globular domain regulates gene expression in yeast. *Cell* **121**, 375–385 (2005).
7. Schneider, J., Bajwa, P., Johnson, F. C., Bhaumik, S. R. & Shilatifard, A. Rtt109 is required for proper H3K56 acetylation: a chromatin mark associated with the elongating RNA polymerase II. *J. Biol. Chem.* **281**, 37270–37274 (2006).
8. Recht, J. *et al.* Histone chaperone Asf1 is required for histone H3 lysine 56 acetylation, a modification associated with S phase in mitosis and meiosis. *Proc. Natl Acad. Sci. USA* **103**, 6988–6993 (2006).
9. Masumoto, H., Hawke, D., Kobayashi, R. & Verreault, A. A role for cell-cycle-regulated histone H3 lysine 56 acetylation in the DNA damage response. *Nature* **436**, 294–298 (2005).
10. Tang, Y. *et al.* Fungal Rtt109 histone acetyltransferase is an unexpected structural homolog of metazoan p300/CBP. *Nature Struct. Mol. Biol.* **15**, 738–745 (2008).
11. Liu, X. *et al.* The structural basis of protein acetylation by the p300/CBP transcriptional coactivator. *Nature* **451**, 846–850 (2008).
12. Florence, B. & McGinnis, W. A genetic screen of the *Drosophila* X chromosome for mutations that modify Deformed function. *Genetics* **150**, 1497–1511 (1998).
13. Marcu, M. G. *et al.* Curcumin is an inhibitor of p300 histone acetyltransferase. *Med. Chem.* **2**, 169–174 (2006).
14. Carre, C., Szymczak, D., Pidoux, J. & Antoniewski, C. The histone H3 acetylase dGcn5 is a key player in *Drosophila melanogaster* metamorphosis. *Mol. Cell. Biol.* **25**, 8228–8238 (2005).
15. Celic, I. *et al.* The sirtuins Hst3 and Hst4p preserve genome integrity by controlling histone H3 lysine 56 deacetylation. *Curr. Biol.* **16**, 1280–1289 (2006).
16. Maas, N. L., Miller, K. M., DeFazio, L. G. & Toczyński, D. P. Cell cycle and checkpoint regulation of histone H3 K56 acetylation by Hst3 and Hst4. *Mol. Cell* **23**, 109–119 (2006).
17. Francis, N. J. & Kingston, R. E. Mechanisms of transcriptional memory. *Nature Rev. Mol. Cell Biol.* **2**, 409–421 (2001).
18. Garcia, B. A. *et al.* Organismal differences in post-translational modifications in histones H3 and H4. *J. Biol. Chem.* **282**, 7641–7655 (2007).
19. Tsubota, T. *et al.* Histone H3–K56 acetylation is catalyzed by histone chaperone-dependent complexes. *Mol. Cell* **25**, 703–712 (2007).
20. Tamburini, B. A., Carson, J. J., Adkins, M. W. & Tyler, J. K. Functional conservation and specialization among eukaryotic anti-silencing function 1 histone chaperones. *Eukaryot. Cell* **4**, 1583–1590 (2005).
21. Scholzen, T. & Gerdes, J. The Ki-67 protein: from the known and the unknown. *J. Cell. Physiol.* **182**, 311–322 (2000).
22. Groth, A. *et al.* Human Asf1 regulates the flow of S phase histones during replicational stress. *Mol. Cell* **17**, 301–311 (2005).
23. Schulz, L. L. & Tyler, J. K. The histone chaperone ASF1 localizes to active DNA replication forks to mediate efficient DNA replication. *FASEB J.* **20**, 488–490 (2006).
24. Sanematsu, F. *et al.* Asf1 is required for viability and chromatin assembly during DNA replication in vertebrate cells. *J. Biol. Chem.* **281**, 13817–13827 (2006).
25. Ait-Si-Ali, S. *et al.* CBP/p300 histone acetyl-transferase activity is important for the G1/S transition. *Oncogene* **19**, 2430–2437 (2000).
26. Yao, T. P. *et al.* Gene dosage-dependent embryonic development and proliferation defects in mice lacking the transcriptional integrator p300. *Cell* **93**, 361–372 (1998).
27. Hasan, S., Hassa, P. O., Imhof, R. & Hottiger, M. O. Transcription coactivator p300 binds PCNA and may have a role in DNA repair synthesis. *Nature* **410**, 387–391 (2001).
28. Ait-Si-Ali, S. *et al.* Histone acetyltransferase activity of CBP is controlled by cycle-dependent kinases and oncoprotein E1A. *Nature* **396**, 184–186 (1998).
29. Meshorer, E. *et al.* Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. *Dev. Cell* **10**, 105–116 (2006).
30. Shechter, D., Dormann, H. L., Allis, C. D. & Hake, S. B. Extraction, purification and analysis of histones. *Nature Protocols* **2**, 1445–1457 (2007).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank J. Kadonaga for *Drosophila* embryos and the Sir2 antibody, A. Mazo for the CBP antisera, and E. Nigg for ASF1A/B antisera. We thank C. Wike and the UC microscope core, and the UCCC flow cytometry core for assistance. This work was supported by funding from the National Institutes of Health (grants GM64475 and CA95641) to J.K.T.

**Author Contributions** C.D. performed all the experiments, and devised some of the experiments. K.H. performed all the mass spectrometry analyses, M.S.L. performed all the pathology and immunohistochemistry analyses, J.K.T. and C.D. wrote the manuscript and J.K.T. guided the research.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to J.K.T. ([jessica.tyler@uchsc.edu](mailto:jessica.tyler@uchsc.edu)).

## METHODS

**Yeast strains.** The 'No Asf1' yeast strain is ROY1170; *MAT $\alpha$  ade2-1 LYS2 leu2-3,112 his3-11 trp1-1 ura3-1 asf1::his5+ TELVIII::URA3 HMRA::ADE2 can1-100* (ref. 31). The 'yAsf1' strain is ROY1172; *MAT $\alpha$  ade2-1 LYS2 leu2-3,112 his3-11 trp1-1 ura3-1 TELVIII::URA3 HMRA::ADE2 can1-100* (ref. 31). The 'dAsf1' is BAT014; *MAT $\alpha$  ade2-1 leu2-3,112 his3-11 trp1-1 ura3-1 TEVII::URA HMRA::ADE2 asf1::dAsf1-13myc KAN<sup>20</sup>*. The 'hAsf1A' strain is BAT016; *MAT $\alpha$  ade2-1 leu2-3,112 his3-11 trp1-1 ura3-1 TEVII::URA HMRA::ADE2 asf1::hAsf1A-13myc KAN<sup>20</sup>*. The 'hAsf1B' strain is KDY006; *MAT $\alpha$  ade2-1 leu2-3,112 his3-11 trp1-1 ura3-1 TEVII::URA HMRA::ADE2 asf1::hAsf1B-13myc KAN<sup>20</sup>*.

**Tissue culture media.** *Drosophila* S2 cells were maintained in Schneider's media supplemented with 10% FBS at 30 °C. HeLa cells were grown in DMEM media supplemented with 10% FBS at 37 °C in 5% CO<sub>2</sub> supply. Unless otherwise stated, curcumin was used at 100  $\mu$ M for 24 h, nicotinamide, nicotinic acid or sodium butyrate were used at 25 mM for 12 h. SHSY-5Y cells were grown in media composed of 50% Ham's F12 and 50% MEM, supplemented with 10% FBS at 37 °C in 5% CO<sub>2</sub> supply. SHSY-5Y cells were treated with 0.02 mM retinoic acid for 3 to 6 days to induce differentiation. MCF7 cells were grown in DMEM supplemented with 10% FBS at 37 °C in 5% CO<sub>2</sub> supply. MCF10A cells were grown in DMEM supplemented with 10% FBS, 0.01 mg ml<sup>-1</sup> insulin, 0.5  $\mu$ g ml<sup>-1</sup> hydrocortisone, 0.02  $\mu$ g ml<sup>-1</sup> EGF and 0.1  $\mu$ g ml<sup>-1</sup> cholera toxin. For trypsinization, 0.05% trypsin-EDTA was used for MCF10A/12A, and 0.25% trypsin-EDTA was used for the rest of the cell lines.

**Antibodies and commercial blots.** *Drosophila* H3K56ac was detected using an H3K56ac-specific antibody from Upstate (catalogue number 07-677) whereas human H3K56ac was detected using a rabbit monoclonal H3K56ac specific antisera from Epitomics (2134-1). Gcn5 was detected with an antibody from Abcam (ab52787), total H3 was detected with an antibody from Abcam (ab1791), H3K9ac was detected with an antibody from Upstate (07-352), human CBP was detected with an antibody from Abcam (ab2832) and human p300 was detected with an antibody from Abcam (ab61217). Asf1 was detected with a previously described antibody<sup>32</sup>. The anti-phospho H2AX antibody was from Upstate (07-164). Commercial tissue blots, having normal and cancerous samples, were procured from G Biosciences (TB56 Set I, II). These blots contain lysates (in a denaturing buffer supplemented with a cocktail of protease inhibitors to minimize proteolytic degradation) extracted from human normal and tumour tissues. Equal amounts of proteins were loaded on a 4–20% gradient denaturing PAGE, followed by transfer in a PVDF membrane.

**Isolation of pellet and supernatant fractions.** NP40-induced extraction of detergent-soluble proteins was performed as described elsewhere<sup>33</sup>. In brief, following treatment with 0.1% or 0.5% NP40 for 10 min, centrifugation at 1,800g for 10 min led to the separation of non-chromatin supernatant and chromatin pellet fractions. The pellet fractions were subsequently purified and analysed by western blotting for the presence of H3K56Ac. Supernatant (non-chromatin) and pellet (chromatin) fractions were resolved following 0.1% NP40 treatment as mentioned elsewhere<sup>34</sup>.

**Isolation of histones from *Drosophila* embryos.** The histone extraction from stage 9–13 *Drosophila* embryos was performed according to standard protocol<sup>35</sup>. In brief, the embryos were homogenized in lysis buffer (15 mM Tris, pH 7.5, 60 mM KCl, 15 mM NaCl, 3 mM EDTA, 0.1 mM EGTA, 0.15 mM spermine, 0.5 mM spermidine, 0.2% NP40, 10 mM NaF and protease inhibitors) supplemented with a cocktail of HDAC inhibitors. The nuclear pellet was subsequently treated with 0.4 M HCl for 1 h for histone extraction. The extracted histones were subsequently TCA precipitated and analysed for the H3K56 acetylation levels.

**Immunoprecipitation.** S2 whole-cell extracts were prepared using RIPA buffer (150 mM NaCl, 1% NP40, 0.5% sodium deoxycholate, 0.1% SDS, 50 mM Tris, pH 8, 10 mM NaF, 0.4 mM EDTA, 10% glycerol and protease inhibitors) supplemented with protease inhibitors. The pre-blocked protein-G–Sepharose-bound fly Asf1 antibody was then incubated with whole-cell extracts. Following extensive washes, the bead-bound protein complexes were analysed by western blotting using fly CBP or H3 antibodies.

**RNA interference in human cells.** siRNA transfections were carried out in HeLa cells using INTERFERin Polyplus (409-10) as per the standard protocol. The different siRNA sequences were chosen based on their success in previous studies or were from Dharmacon. CAF-1 (p150)-1, 5'-AGGGGAAAGCCGAUGACAU (dTdT)-3' (ref. 36). ASF1A, 5'-AAGUGAAGAAUACGAUCAAGU(dTdT)-3' (ref. 22). ASF1B, 5'-AACAGACGUAACCUCAACCCU(dTdT)-3' (ref. 22). SIRT1, 5'-ACUUUGCUGUACCCUGUA(dTdT)-3' (ref. 37). SIRT2, ON-TARGETplus SMARTpool L-004826 (from Dharmacon). CBP, 5'-CGGCACAG CCUCUCAGUCA(dTdT)-3' (ref. 38). p300, 5'-UGACACAGGCAGGCUUG ACUU-3 (ref. 39).

For CAF-1 and ASF1A/B, the siRNA transfection time was for 48 h. For CBP, p300, SIRT1 and SIRT2, the transfection time was 72 h.

**Flow cytometry.** After the different treatments, the cells were washed with PBS and subject to ethanol fixation for at least 1 h. The cells were then Triton-extracted (0.05%), blocked in 1% BSA, and stained with H3K56ac antibody (Epitomics) at 1:250 dilution for 1 h. After extensive washing in PBST, the cells were put in Alexa-568-conjugated secondary antibody for 1 h at 1:1,000 dilution. The cells were extensively washed, 0.5 mg ml<sup>-1</sup> RNase treated and finally stained with propidium iodide at 10  $\mu$ g ml<sup>-1</sup> final concentration for 1 h in the dark. The cell cycle stages and K56ac staining intensities were then measured by flow cytometry.

**Immunohistochemistry.** After deparafinizing, slides were rinsed in ethanol. Antigen retrieval was performed with 10 mM sodium citrate for 5 min at 22 psi. The staining was performed with the I-VIEW enhanced DAB kit. Counterstaining was performed with Mayer's haematoxylin, and the mounting media used was Cytoseal 60. IMH-346 and IMT-01233 tissue array slides were procured from Imgenex. Additional slides for normal and tumour skin, thyroid, cerebellum, colon, larynx and ovary tissues were analysed from the UCCC Pathology core. Each of the sample specimens was pathologically confirmed before carrying out the immunohistochemistry.

**DNA-damaging treatments.** Unless otherwise indicated, 0.05% MMS and 150 mM hydroxyurea treatment was performed for 12 h. 49,995 J m<sup>-2</sup> of ultra-violet or 1, 1.25 and 1.5 Gy gamma irradiation were given to visualize the damage foci.

**Immunofluorescence.** In brief, following 4% formaldehyde fixation, cells were permeabilized by 0.5% Triton X-100, blocked in 1% BSA, and treated with primary antibody followed by Alexa-488-conjugated secondary antibody. Vectashield mounting medium containing DAPI was used for DNA staining. H3K56ac and H2AXp antibodies were detected using Alexa-488- or Alexa-568-conjugated secondary antibody, respectively, following previously described methods<sup>32</sup>.

**Mass spectrometry analysis.** The histone K56 acetylation analysis by mass spectrometry sample digestion A standard in-gel digestion protocol was used based on previously used methods<sup>40,41</sup>. Iodoacetamide (IAM) was used for Cys alkylation. Liquid chromatography–tandem mass spectrometry data acquisition digests were analysed on a LTQ-ICR hybrid mass spectrometer (LTQ-FT Ultra, Thermo Fisher). Peptide desalting and separation was achieved using a dual capillary/nano pump HPLC system (Agilent 1200). On this system 8  $\mu$ l of sample was loaded onto a trapping column (ZORBAX 300SB-C18, dimensions 5  $\times$  0.3 mm (5  $\mu$ m); Agilent Technologies) and washed with 5% acetonitrile, 0.1% formic acid at a flow rate of 15  $\mu$ l min<sup>-1</sup> for 5 min. At this time the trapping column was put online with the nano-pump at a flow rate of 350 nl min<sup>-1</sup>. An 80 min gradient from 8% ACN to 32% ACN was used to separate the peptides. The column was made from an in-house pulled 360/100 nm (outer/inner diameter) fused silica capillary using a model P-2000 laser puller (Sutter Instrument Co.). The column was packed 15 cm in length with Jupiter C18 resin (Phenomenex). The column was kept at a constant 40 °C using an in-house built column heater. The column effluent was coupled directly to the mass spectrometer with an in-house built nanospray ion source. Data acquisition was performed using the instrument supplied Xcalibur (version 2.0.7) software. The 90-min liquid chromatography runs were monitored by sequentially recording the precursor scan (MS) followed by three collision-induced dissociation (CID) acquisitions (MS/MS). Normalized collision energies were employed using helium as the collision gas. MS survey scans were acquired in the ICR cell at a resolution of 25,000 at 400 *m/z*. After two acquisitions of a given ion within 45 s, the ion was excluded for 150 s.

**Data analysis.** The manufacturer-supplied extract\_msn script was used to create de-isotoped, centroided peak lists from the raw spectra (.mgf format). These peak lists were searched against the SwissProt database (v54.8) using Mascot server (Version 2.2, Matrix Science). For searches, mass tolerances were  $\pm$  10 parts per million (p.p.m.) for MS peaks, and  $\pm$  0.6 Da for MS/MS fragment ions. Trypsin specificity was used allowing for one missed cleavage. The modifications of Met oxidation, protein N-terminal acetylation, peptide N-terminal pyroglutamic acid formation (Q), lysine acetylation and lysine mono-, di- and tri-methylation were allowed for. Peptides were required to reach a score of 6 and proteins were filtered at a significance of *P* < 0.005 and required bold red hits (top match for a given spectrum).

**Semiquantification of K56 acetylation.** The tryptic peptide resulting from the unacetylated K56 form of H3 was 'K.STELLIR.K'. This peptide was compared to the tryptic peptide resulting from the K56 acetylated form of H3 'R.YQ-AcK-STELLIR.K'. Trypsin cleavage does not occur on the C-terminal side of acetylated K56 due to the loss of the positive charge. Extracted ion chromatograms were calculated for the range *m/z* 416.23–416.27 for the 'STELLIR' peptide. It was confirmed that the peak from ~27 to 30 min corresponded to the 'STELLIR' peptide, based on assignment of several CID spectra. Likewise the extracted ion chromatogram from 646.81 to 646.91 was used to calculate the acetylated form



of the peptide. The identification of this peak was also confirmed by several redundant CID spectra. Peak areas were calculated using the 'Genesis' Algorithm from within the Bioworks Qual Browser (version 2.0, Thermo Fisher). The data are stated as the percentage of K56ac with respect to total K56 and K56Ac. Although this isn't a direct measure of percentage acetylation, it does show that K56ac increases and accordingly that K56 decreases. Note that our analyses do not distinguish among the three different histone variants: H3.3 versus H3.1 versus H3.2.

**Verification of H3K56 acetylation rather than H3K56 trimethylation.** Note that we can readily distinguish the acetylated lysine 56 and trimethylated lysine 56 by mass using our analytical platform. The mass difference between acetylation ( $\text{H}_2\text{C}_2\text{O}$ , 42.01056 Da) and tri-methylation ( $\text{H}_6\text{C}_3$ , 42.04695 Da) of the peptide in question can readily be distinguished using the FT-ICR instrument where we routinely measure ions at less than 1 p.p.m. mass accuracy. The average parent ion error for H3 was  $-0.66$  p.p.m., and  $3 \times \text{StdDev}$  ( $3\sigma$ ) of all the assigned peptides was 1.26 p.p.m.. The mass error for the ion used to assign YQK(Ac)STELLIR from the *in vitro* derived sample had a mass error of  $-0.74$  p.p.m. whereas the mass error for the tri-methylated form would have been  $-28.9$  p.p.m.—well outside the acceptable error tolerance for the mass spectrometer used and observed for the rest of the peptide assignments. Similarly, the mass error for assignment of the peptide YQK(Ac)STELLIR from the *in vivo* samples was  $-1.7$  p.p.m., and the error if this precursor would be assigned as the tri-methylated form would be an unacceptable  $-29.8$  p.p.m.

31. Tyler, J. K. *et al.* The RCAF complex mediates chromatin assembly during DNA replication and repair. *Nature* **402**, 555–560 (1999).
32. Tyler, J. K. *et al.* Interaction between the *Drosophila* CAF-1 and ASF1 chromatin assembly factors. *Mol. Biol. Cell* **21**, 6574–6584 (2001).
33. Frei, C. & Gasser, S. M. The yeast Sgs1p helicase acts upstream of Rad53p in the DNA replication checkpoint and colocalizes with Rad53p in S-phase-specific foci. *Genes Dev.* **14**, 81–96 (2000).
34. Pallier, C. *et al.* Association of chromatin proteins high mobility group box (HMGB) 1 and HMGB2 with mitotic chromosomes. *Mol. Biol. Cell* **14**, 3414–3426 (2003).
35. Madigan, J. P., Chotkowski, H. L. & Glaser, R. L. DNA double-strand break-induced phosphorylation of *Drosophila* histone variant H2Av helps prevent radiation-induced apoptosis. *Nucleic Acids Res.* **30**, 3698–3705 (2002).
36. Hoek, M. & Stillman, B. Chromatin assembly factor 1 is essential and couples chromatin assembly to DNA replication *in vivo*. *Proc. Natl Acad. Sci. USA* **100**, 12183–12188 (2003).
37. Kim, E. J., Kho, J. H., Kang, M. R. & Um, S. J. Active regulator of SIRT1 cooperates with SIRT1 and facilitates suppression of p53 activity. *Mol. Cell* **28**, 277–290 (2007).
38. Megiorni, F., Indovina, P., Mora, B. & Mazzilli, M. C. Minor expression of fascin-1 gene (FSCN1) in Ntera2 cells depleted of CREB-binding protein. *Neurosci. Lett.* **381**, 169–174 (2005).
39. Kuninger, D. *et al.* Gene disruption by regulated short interfering RNA expression, using a two-adenovirus system. *Hum. Gene Ther.* **15**, 1287–1292 (2004).
40. Rosenfeld, J., Capdevielle, J., Guillemot, J. C. & Ferrara, P. In-gel digestion of proteins for internal sequence analysis after one- or two-dimensional gel electrophoresis. *Anal. Biochem.* **203**, 173–179 (1992).
41. Hellman, U., Wernstedt, C., Genez, J. & Heldin, C. H. Improvement of an "In-Gel" digestion procedure for the micropreparation of internal protein fragments for amino acid sequencing. *Anal. Biochem.* **224**, 451–455 (1995).

## LETTERS

# Hypusine-containing protein eIF5A promotes translation elongation

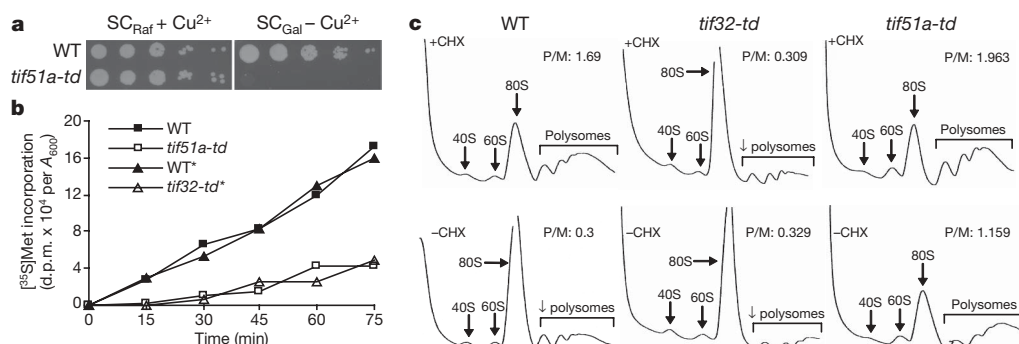
Preeti Saini<sup>1</sup>, Daniel E. Eyler<sup>2</sup>, Rachel Green<sup>2</sup> & Thomas E. Dever<sup>1</sup>

Translation elongation factors facilitate protein synthesis by the ribosome. Previous studies identified two universally conserved translation elongation factors, EF-Tu in bacteria (known as eEF1A in eukaryotes) and EF-G (eEF2), which deliver aminoacyl-tRNAs to the ribosome and promote ribosomal translocation, respectively<sup>1</sup>. The factor eIF5A (encoded by *HYP2* and *ANB1* in *Saccharomyces cerevisiae*), the sole protein in eukaryotes and archaea to contain the unusual amino acid hypusine (*N*<sup>ε</sup>-(4-amino-2-hydroxybutyl)lysine)<sup>2</sup>, was originally identified based on its ability to stimulate the yield (endpoint) of methionyl-puromycin synthesis—a model assay for first peptide bond synthesis thought to report on certain aspects of translation initiation<sup>3,4</sup>. Hypusine is required for eIF5A to associate with ribosomes<sup>5,6</sup> and to stimulate methionyl-puromycin synthesis<sup>7</sup>. Because eIF5A did not stimulate earlier steps of translation initiation<sup>8</sup>, and depletion of eIF5A in yeast only modestly impaired protein synthesis<sup>9</sup>, it was proposed that eIF5A function was limited to stimulating synthesis of the first peptide bond or that eIF5A functioned on only a subset of cellular messenger RNAs. However, the precise cellular role of eIF5A is unknown, and the protein has also been linked to mRNA decay, including the non-sense-mediated mRNA decay pathway<sup>10,11</sup>, and to nucleocytoplasmic transport<sup>12,13</sup>. Here we use molecular genetic and biochemical studies to show that eIF5A promotes translation elongation. Depletion or inactivation of eIF5A in the yeast *S. cerevisiae* resulted in the accumulation of polysomes and an increase in ribosomal transit times. Addition of recombinant eIF5A from yeast, but not

a derivative lacking hypusine, enhanced the rate of tripeptide synthesis *in vitro*. Moreover, inactivation of eIF5A mimicked the effects of the eEF2 inhibitor sordarin, indicating that eIF5A might function together with eEF2 to promote ribosomal translocation. Because eIF5A is a structural homologue of the bacterial protein EF-P<sup>14,15</sup>, we propose that eIF5A/EF-P is a universally conserved translation elongation factor.

To help clarify the function of eIF5A, encoded by the *TIF51A* (*HYP2*) and *TIF51B* (*ANB1*) genes in *S. cerevisiae*, we deleted the non-essential *TIF51B* gene and constructed a copper-regulated *TIF51A* degon mutant (*tif51a-td*) that produces a ubiquitin–Arg-codon–eIF5A fusion protein (UBI–R–eIF5A). Under permissive conditions the *tif51a-td* mutant grew similarly to the isogenic wild-type strain (Fig. 1a and Supplementary Fig. 1a). After a shift to non-permissive conditions, UBI–R–eIF5A was depleted by 6 h and the growth rate of the *tif51a-td* strain was substantially reduced (Fig. 1a and Supplementary Fig. 1). This growth defect was complemented by a wild-type *TIF51A* plasmid (data not shown). As shown in Fig. 1b, depletion of eIF5A after 14 h growth of the *tif51a-td* mutant under non-permissive conditions, like depletion of the eIF3a (encoded by *TIF32*, also known as *RPG1*) subunit of the translation initiation factor eIF3 (ref. 16), caused a severe reduction in [<sup>35</sup>S]methionine incorporation into total protein.

Polyribosome profiles of whole-cell extracts (WCEs) from wild-type and *tif51a-td* strains grown under nonpermissive conditions were analysed by velocity sedimentation in sucrose gradients. In wild-type cells treated with cycloheximide (CHX) to inhibit translation



**Figure 1 | eIF5A depletion impairs yeast cell growth and protein synthesis, and causes retention of polysomes.** **a**, Isogenic wild-type (WT, J714) and *tif51a-td* mutant (J702) strains were serially diluted, spotted on permissive (SC<sub>Raf</sub> + Cu<sup>2+</sup>) or non-permissive (SC<sub>Gal</sub>) medium, and incubated 5 days at 25 °C. Total protein synthesis (**b**) and polysome profiles (**c**) were analysed in *tif51a-td* (J702), *tif32-td* (YAJ22) and their isogenic wild-type strains after growth under non-permissive conditions. Incorporation (disintegrations per min, d.p.m.) of [<sup>35</sup>S]Met is expressed per A<sub>600</sub> unit, and results are

representative of triplicate experiments. WT\* and *tif32-td\** strains are isogenic. **c**, Cells were treated with 50 μg ml<sup>-1</sup> cycloheximide (CHX) before collection (+) or were left untreated (–), and WCEs were separated on sucrose gradients and fractionated to visualize the indicated ribosomal species. P/M ratios were calculated by comparing areas under the 80S and polysome peaks; the ‘down arrow polysomes’ label indicates a reduced amount of polysomes (due to run-off).

<sup>1</sup>Laboratory of Gene Regulation and Development, NICHD, National Institutes of Health, Bethesda, Maryland 20892, USA. <sup>2</sup>Howard Hughes Medical Institute, Department of Molecular Biology and Genetics, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA.

elongation, distinct 40S and 60S ribosomal subunit peaks, as well as 80S monosome and polysome peaks, were detected (Fig. 1c, top left panel). Mutations that impair translation initiation, such as the eIF3a degon, reduce the amount of polysomes and cause a corresponding increase in the 80S peak, which consists of both 80S monosomes translating an mRNA and 80S complexes not engaged with an mRNA (note the fivefold reduction in polysome to monosome (P/M) ratio in the *tif32-td* versus wild-type strain; Fig. 1c, top left and middle panels). In contrast, inhibition of translation in the *tif51a-td* strain resulted in a polysome profile indistinguishable from that of the wild-type strain and a similar P/M ratio (Fig. 1c, upper panels). Thus, depletion of eIF5A did not appreciably impair translation initiation.

To monitor post-initiation defects, polysome analyses were performed in the absence of CHX. Under these conditions, ribosomes in a wild-type strain continue elongating during extract preparation, run-off the mRNA (complete peptide synthesis and disengagement from the mRNA) and accumulate as vacant 80S couples (Fig. 1c, bottom left panel). In contrast, a defect in translation elongation or termination results in slower ribosome run-off and the retention of polysomes, as measured by an increase in the P/M ratio. As shown in Fig. 1c (bottom panels), polysomes were retained in the *tif51a-td* strain in the absence of CHX resulting in a fourfold increase in the P/M ratio—that is, mimicking the effect of CHX on the wild-type strain. If eIF5A only functioned during synthesis of the first peptide bond, as was previously proposed<sup>8</sup>, then post-initiation ribosomes would be expected to continue elongating and run-off the mRNA during extract preparation (like the *tif32-td* initiation mutant in Fig. 1c). The stable polysomes in the *tif51a-td* strain, which were observed in crude cell extracts and thus include most of the mRNAs in the cell, indicate a general translation elongation/termination defect in the absence of eIF5A.

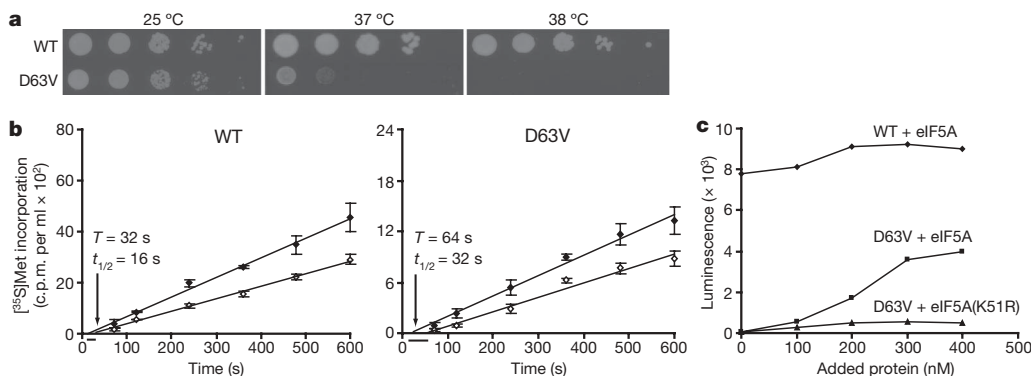
To inactivate eIF5A rapidly, we identified a new temperature-sensitive mutation in *TIF51A*. The *tif51a(D63V)* mutant exhibited a slow-growth phenotype at permissive temperature (25 °C), which was exacerbated at 37 °C and lethal at 38 °C (Fig. 2a), consistent with substantial loss of the protein *in vivo* (data not shown). Both the *tif51a(D63V)* mutant and the previously characterized *tif51a(S149P)* mutant, isolated as a suppressor of nonsense-mediated mRNA decay (NMD)<sup>11</sup>, stabilized polysomes when incubated for 2 h at the non-permissive temperature (Supplementary Fig. 2). Whereas mutations that impair elongation or termination could potentially cause polysome retention in the absence of CHX, our analysis (Supplementary Fig. 3a) and previously published results<sup>17</sup> of yeast termination factor eRF3 (encoded by *SUP35*, also known as *SUP2*) mutants revealed polysome run-off. In contrast, the *tif51a(D63V)* and *tif51a(S149P)*

mutants (Supplementary Fig. 2), like elongation factor eEF3 (*YEF3*) and eEF2 (*EFT2*) mutants<sup>18,19</sup> and like the eEF2 inhibitor sordarin (Supplementary Fig. 3b), caused polysome accumulation. Thus, the primary defect *in vivo* upon inactivation of eIF5A is impaired translation elongation.

To provide an independent assessment of translation elongation/termination, we determined ribosomal transit times in wild-type and *tif51a(D63V)* mutant strains cultured at a semi-permissive temperature (36 °C). The ribosomal transit time is the time required for a ribosome, after initiation, to decode an mRNA and release the completed polypeptide chain<sup>20,21</sup>. Whereas [<sup>35</sup>S]Met incorporation into total protein (nascent chains plus completed polypeptides) increases linearly upon addition of the labelled amino acid to cells, there is a lag before [<sup>35</sup>S]Met incorporation into completed (disengaged from the ribosome and therefore soluble) proteins increases linearly. Plotting the incorporation of [<sup>35</sup>S]Met into total (nascent plus completed) and completed proteins as a function of time yields two lines. The displacement (lag) between the lines reflects the time required to uniformly label the nascent polypeptides; the average transit time is calculated by doubling the difference in the *x*-axis intercepts for the two lines. The average ribosomal transit time in the *tif51a(D63V)* mutant strain at 36 °C (~64 s) was ~2-fold slower than that in the wild-type strain (~32 s; Fig. 2b), confirming that inactivation of eIF5A impairs translation elongation/termination.

The translation defect in the eIF5A mutant strains could be due to a direct role of eIF5A in translation or an indirect effect of eIF5A on the translational machinery. After a brief (5 min) incubation at the restrictive temperature (38 °C), extracts from the *tif51a(D63V)* mutant were inactive in translating a luciferase reporter mRNA (Fig. 2c, *y*-axis intercept). Addition of recombinant wild-type eIF5A purified from yeast, but not unhyposinated eIF5A(K51R), resulted in a dose-dependent restoration of translational activity (Fig. 2c). Thus, eIF5A seems to stimulate translation directly and to function as a general translation elongation factor in a manner dependent on its hypusine modification.

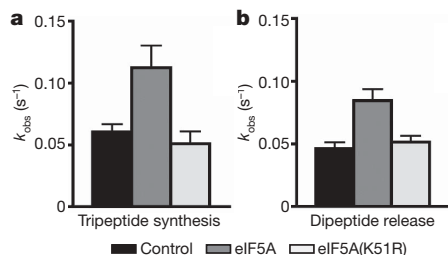
To extend the *in vitro* studies, we examined the effects of eIF5A using an *in vitro* reconstituted translation system. First, we confirmed previous reports<sup>7,8</sup> that eIF5A, but not unhyposinated eIF5A(K51R), stimulated methionyl-puromycin synthesis—a model assay of first peptide bond synthesis (data not shown). Next, yeast ribosomes were programmed with a defined mRNA encoding Met-Phe-Phe-Stop (AUGUUCUUCUAA) and initiator [<sup>35</sup>S]Met-tRNA<sub>i</sub><sup>Met</sup> using purified initiation factors (eIF1, eIF1A, eIF2, eIF5 and eIF5B)<sup>22</sup>, and these initiation complexes were pelleted over a sucrose cushion. To assay elongation, initiation complexes were reacted with an elongation factor



**Figure 2 | Translation elongation defect in temperature-sensitive *tif51a(D63V)* mutant.** **a**, Isogenic wild-type (WT, J697) and *tif51a(D63V)* mutant (D63V, J698) strains were serially diluted, spotted on SC medium, and incubated 3 days at 25, 37 or 38 °C. **b**, Wild-type (left panel) and *tif51a(D63V)* (right) mutant strains were shifted to 36 °C for 2 h, and then labelled with [<sup>35</sup>S]Met. Fitting lines of [<sup>35</sup>S]Met incorporation (counts per minute, c.p.m., per ml) into total (filled symbols) and completed (open

symbols) protein synthesis were obtained by linear regression and used to determine the half-average (*t*<sub>1/2</sub>) and average (*T*) transit time. Error bars are the standard deviation from three independent experiments. **c**, *In vitro* translation activity of heat-treated extracts from wild-type and *tif51a(D63V)* strains following addition of eIF5A or eIF5A(K51R). Results are representative of six independent experiments.





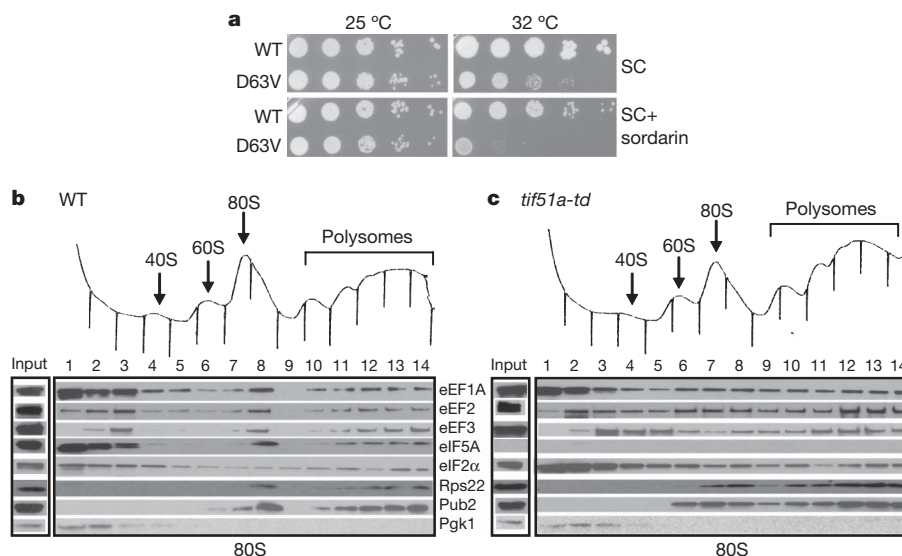
**Figure 3 | eIF5A stimulates elongation and termination *in vitro*.** Rate constants for tripeptide synthesis (a) and dipeptide release (b) in the presence of eIF5A, eIF5A(K51R) or no added factors. Error bars are the standard error from three independent experiments.

mix containing eEF1A–Phe–tRNA<sup>Phe</sup>–GTP ternary complex, eEF2, eEF3, GTP and ATP along with eIF5A or the unhyposinated eIF5A(K51R) mutant. The rate of production of tripeptide product (Met–Phe–Phe) was stimulated approximately twofold by eIF5A in this purified system whereas the variant eIF5A(K51R) had no apparent effect (Fig. 3a and Supplementary Fig. 4a). We also examined the rate of peptide release (termination) using pelleted elongation complexes (carrying an mRNA with a dipeptide open reading frame (ORF) and a dipeptidyl–tRNA ([<sup>35</sup>S]Met–Phe–tRNA<sup>Phe</sup>) in the P site) to which purified eRF1 and eRF3 were added and the products of the reaction analysed in a similar fashion. As for the elongation steps, the rate of the termination reaction was stimulated approximately twofold by wild-type eIF5A, but not by the unhyposinated variant (eIF5A(K51R); Fig. 3b and Supplementary Fig. 4b). In both cases (elongation or termination) there were no effects on the extent (endpoint) of the reaction. Taken together, these data indicate that eIF5A optimizes the ribosome complex for more productive interactions with the mechanistically related tRNA and release factor substrates<sup>23</sup>.

Consistent with eIF5A functioning in translation elongation, the *tif51a*(D63V) and *tif51a*(S149P) mutants impaired programmed ribosomal frameshifting. The eIF5A mutants, like treatment of the wild-type strain with the eEF2 inhibitor sordarin, caused a similar and specific inhibition of +1, but not –1, frameshifting (Supplementary Fig. 5, see also ref. 24). Because sordarin also exacerbated the growth defect of the *tif51a*(D63V) mutant (Fig. 4a), these data indicate that eIF5A and eEF2 functionally interact in the ribosome elongation

cycle. In accordance, eIF5A associated with polysomes *in vivo*. After treatment with formaldehyde to freeze polysomes and cross-link factors to the ribosome<sup>16</sup>, WCEs were subjected to velocity sedimentation in sucrose gradients. As shown in Fig. 4b, eIF5A, like elongation factors eEF1A, eEF2 and eEF3, was found both at the top of the gradient and in 80S complexes and polysomes. In contrast, 3-phosphoglycerate kinase (Pgk1), a soluble cytoplasmic protein, was found exclusively at the top of the gradient (Fig. 4b), the peripheral ribosomal 40S protein Asc1 (yeast orthologue of Rack1) co-sedimented with the integral 40S protein Rps22 (data not shown), and initiation factors eIF2 (Fig. 4b) and eIF3 (data not shown) were distributed across the gradient and did not accumulate on 80S complexes like eIF5A and the elongation factors. As shown in Fig. 4c, depletion of eIF5A in the *tif51a*-td strain resulted in an ~2-fold accumulation of elongation factor eEF2 on polysomes. These results, combined with the ability of glutathione S-transferase (GST)–eIF5A to pull-down eEF2 and 80S ribosomes from crude yeast extracts<sup>6</sup>, support the notion that eIF5A functionally interacts with eEF2 to promote translation elongation.

Owing to its unique and essential hypusine modification, eIF5A has been of considerable interest. Our results provide evidence that eIF5A is a translation elongation factor and provide a rationale for its previously proposed roles in translation initiation and mRNA turnover. It is noteworthy that the putative translation initiation function of eIF5A was assigned based on its ability to promote methionyl-puromycin synthesis<sup>3,4,8</sup>, essentially a translation elongation step. Similarly, because NMD requires ongoing translation, the inhibition of NMD by a mutation in eIF5A<sup>11</sup> is consistent with the observation that CHX blocks this form of mRNA turnover<sup>25</sup>. eIF5A is a structural homologue of the bacterial protein EF–P<sup>14,15</sup>, which was reported to enhance methionyl-puromycin and poly-Phe synthesis in reconstituted translation assays<sup>26,27</sup>. Because both EF–P and eIF5A alter the Mg<sup>2+</sup> optimum for protein synthesis *in vitro*<sup>4,27</sup>, the failure to detect the critical role of eIF5A/EF–P in translation elongation previously might stem from the fact that early biochemical screens were performed using *in vitro* translation reactions with abundant Mg<sup>2+</sup> levels. Taken together, our studies reveal that, in addition to eEF1A/EF–Tu and eEF2/EF–G, a third universal and essential factor, eIF5A/EF–P, is required for translation elongation. Future studies will seek to determine at a molecular level how eIF5A promotes the ribosomal reactions required for translation elongation.



**Figure 4 | Functional connection between eIF5A and eEF2.** a, Isogenic wild-type (WT, J697) and *tif51a*(D63V) (D63V, J698) strains were serially diluted, spotted on SC medium lacking or containing 200 ng ml<sup>–1</sup> sordarin, and incubated 3 days at 25 or 32 °C. b, c, Cultures of wild-type or *tif51a*-td strains

grown under non-permissive conditions were cross-linked with 1% formaldehyde. WCEs were separated on sucrose gradients, and gradient fractions were subjected to western blot analysis using antibodies against the indicated proteins (Pub2, 60S subunit protein Rpl39).

## METHODS SUMMARY

The *tif51a*-*td* degon allele was constructed using methods described previously<sup>16</sup> by inserting a DNA cassette carrying a copper-regulated promoter, the protein-destabilizing ubiquitin-coding region and an Arg codon (pCUP1-UBI-R)<sup>28</sup> upstream of and in-frame with the *TIF51A* ORF. To enhance degradation of the UBI-R-eIF5A fusion protein, a galactose-inducible version of the ubiquitin E3-ligase Ubr1 (ref. 29) was introduced into the *tif51a*-*td* strain. The *tif51a* temperature-sensitive mutants were isolated and growth assays of tenfold serial dilutions of yeast strains were performed using standard methods. Degon strains were grown under permissive (synthetic complete (SC) medium containing 2% raffinose and 100  $\mu$ M copper sulphate (SC<sub>Raf</sub> + Cu<sup>2+</sup>), where the *CUP1* promoter is induced and *UBR1* expression is low) or under non-permissive (SC medium containing 2% galactose lacking copper (SC<sub>Gal</sub>), which represses new synthesis and triggers proteasomal degradation of UBI-R-eIF5A) conditions. Plasmids and yeast strains are listed in Supplementary Table 1 and Supplementary Table 2, respectively. Yeast polysome analyses from formaldehyde cross-linked cells and analysis of ribosomal complexes were performed as described previously<sup>16</sup>. Anti-yeast eIF5A antiserum was obtained from rabbits immunized with recombinant N-terminal poly-histidine-tagged eIF5A produced in bacteria. The assays for programmed ribosomal frameshifting<sup>30</sup> were conducted as described previously. The ribosomal transit time measurements<sup>21</sup> and the reconstituted translation elongation assays<sup>22</sup> were extensions of previously established methods.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 4 November 2008; accepted 23 March 2009.**

- Merrick, W. C. & Nyborg, J. in *Translational Control of Gene Expression* (eds Sonenberg, N., Hershey, J. W. B. & Mathews, M. B.) 89–125 (Cold Spring Harbor Laboratory Press, 2000).
- Wolff, E. C., Kang, K. R., Kim, Y. S. & Park, M. H. Posttranslational synthesis of hypusine: evolutionary progression and specificity of the hypusine modification. *Amino Acids* **33**, 341–350 (2007).
- Kemper, W. M., Berry, K. W. & Merrick, W. C. Purification and properties of rabbit reticulocyte protein synthesis initiation factors M2B $\alpha$  and M2B $\beta$ . *J. Biol. Chem.* **251**, 5551–5557 (1976).
- Schreier, M. H., Erni, B. & Staehelin, T. Initiation of mammalian protein synthesis: purification and characterization of seven initiation factors. *J. Mol. Biol.* **116**, 727–753 (1977).
- Jao, D. L. & Chen, K. Y. Tandem affinity purification revealed the hypusine-dependent binding of eukaryotic initiation factor 5A to the translating 80S ribosomal complex. *J. Cell. Biochem.* **97**, 583–598 (2006).
- Zanelli, C. F. *et al.* eIF5A binds to translational machinery components and affects translation in yeast. *Biochem. Biophys. Res. Commun.* **348**, 1358–1366 (2006).
- Park, M. H., Wolff, E. C., Smit-McBride, Z., Hershey, J. W. & Folk, J. E. Comparison of the activities of variant forms of eIF-4D. The requirement for hypusine or deoxyhypusine. *J. Biol. Chem.* **266**, 7988–7994 (1991).
- Benne, R. & Hershey, J. W. B. The mechanism of action of protein synthesis initiation factors from rabbit reticulocytes. *J. Biol. Chem.* **253**, 3078–3087 (1978).
- Kang, H. A. & Hershey, J. W. Effect of initiation factor eIF-5A depletion on protein synthesis and proliferation of *Saccharomyces cerevisiae*. *J. Biol. Chem.* **269**, 3934–3940 (1994).
- Schrader, R., Young, C., Kozian, D., Hoffmann, R. & Lottspeich, F. Temperature-sensitive eIF5A mutant accumulates transcripts targeted to the nonsense-mediated decay pathway. *J. Biol. Chem.* **281**, 35336–35346 (2006).
- Zuk, D. & Jacobson, A. A single amino acid substitution in yeast eIF-5A results in mRNA stabilization. *EMBO J.* **17**, 2914–2925 (1998).
- Ruhl, M. *et al.* Eukaryotic initiation factor 5A is a cellular target of the human immunodeficiency virus type 1 Rev activation domain mediating trans-activation. *J. Cell Biol.* **123**, 1309–1320 (1993).
- Zanelli, C. F. & Valentini, S. R. Is there a role for eIF5A in translation? *Amino Acids* **33**, 351–358 (2007).
- Hanawa-Suetsugu, K. *et al.* Crystal structure of elongation factor P from *Thermus thermophilus* HB8. *Proc. Natl Acad. Sci. USA* **101**, 9595–9600 (2004).
- Kyrpides, N. C. & Woese, C. R. Universally conserved translation initiation factors. *Proc. Natl Acad. Sci. USA* **95**, 224–228 (1998).
- Jivotovskaya, A. V., Valasek, L., Hinnebusch, A. G. & Nielsen, K. H. Eukaryotic translation initiation factor 3 (eIF3) and eIF2 can promote mRNA binding to 40S subunits independently of eIF4G in yeast. *Mol. Cell. Biol.* **26**, 1355–1372 (2006).
- Smirnov, V. N. *et al.* Recessive nonsense-suppression in yeast: further characterization of a defect in translation. *FEBS Lett.* **66**, 12–15 (1976).
- Anand, M., Chakraborty, K., Marton, M. J., Hinnebusch, A. G. & Kinzy, T. G. Functional interactions between yeast translation eukaryotic elongation factor (eEF) 1A and eEF3. *J. Biol. Chem.* **278**, 6985–6991 (2003).
- Ortiz, P. A. & Kinzy, T. G. Dominant-negative mutant phenotypes and the regulation of translation elongation factor 2 levels in yeast. *Nucleic Acids Res.* **33**, 5740–5748 (2005).
- Fan, H. & Penman, S. Regulation of protein synthesis in mammalian cells. II. Inhibition of protein synthesis at the level of initiation during mitosis. *J. Mol. Biol.* **50**, 655–670 (1970).
- Nielsen, P. J. & McConkey, E. H. Evidence for control of protein synthesis in HeLa cells via the elongation rate. *J. Cell. Physiol.* **104**, 269–281 (1980).
- Acker, M. G., Koltz, S. E., Mitchell, S. F., Nanda, J. S. & Lorsch, J. R. Reconstitution of yeast translation initiation. *Methods Enzymol.* **430**, 111–145 (2007).
- Youngman, E. M., McDonald, M. E. & Green, R. Peptide release on the ribosome: mechanism and implications for translational control. *Annu. Rev. Microbiol.* **62**, 353–373 (2008).
- Harger, J. W., Meskauskas, A., Nielsen, J., Justice, M. C. & Dinman, J. D. Ty1 retrotransposition and programmed +1 ribosomal frameshifting require the integrity of the protein synthetic translocation step. *Virology* **286**, 216–224 (2001).
- Zhang, S. *et al.* Polysome-associated mRNAs are substrates for the nonsense-mediated mRNA decay pathway in *Saccharomyces cerevisiae*. *RNA* **3**, 234–244 (1997).
- Glick, B. R. & Ganoza, M. C. Identification of a soluble protein that stimulates peptide bond synthesis. *Proc. Natl Acad. Sci. USA* **72**, 4257–4260 (1975).
- Ganoza, M. C. & Aoki, H. Peptide bond synthesis: function of the *efp* gene product. *Biol. Chem.* **381**, 553–559 (2000).
- Dohmen, R. J., Wu, P. & Varshavsky, A. Heat-inducible degon: a method for constructing temperature-sensitive mutants. *Science* **263**, 1273–1276 (1994).
- Labib, K., Tercero, J. A. & Diffley, J. F. Uninterrupted MCM2–7 function required for DNA replication fork progression. *Science* **288**, 1643–1647 (2000).
- Harger, J. W. & Dinman, J. D. An *in vivo* dual-luciferase assay system for studying translational recoding in the yeast *Saccharomyces cerevisiae*. *RNA* **9**, 1019–1024 (2003).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank T. G. Kinzy for providing anti-yeast eEF1A, eEF2 and eEF3 antisera, as well as constructs for the purification of eEF2 and eEF3; J. Lorsch for providing constructs for purification of yeast initiation factors; J. Dinman for frameshifting reporter vectors, and A. Hinnebusch, T. G. Kinzy, A. Jivotovskaya, D. Shelton, C. Grant, J. Lorsch and members of the Dever and Hinnebusch laboratories for comments and discussion. Salary support provided by HHMI (R.G.) and NIH (D.E.E.). This work was supported in part by the Intramural Research Program of the NIH, NICHD (to T.E.D.).

**Author Contributions** P.S., D.E.E., R.G. and T.E.D. designed the experiments and wrote the manuscript. P.S. performed all experiments except reconstituted translation experiments, which were performed by D.E.E.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to T.E.D. (tdever@nih.gov).

## METHODS

**Plasmid and yeast strain construction.** A *tif51a*-*td* integration plasmid (pC3286) was constructed by inserting an XhoI-flanked PCR (primers: P1, 5'-CGCTCGAGATGTCTGACGAGAACATACCTTTGAA-3', and P2, 5'-CCGC TCGAGGTGAGTAGATGGAGACAAATCTTCCAAC-3') product containing the first 225 bp of the eIF5A ORF into the XhoI site of p4437, a plasmid containing the copper-regulated *CUP1* promoter upstream of the ubiquitin coding region followed by an Arg codon (pCUP1-UBI-R)<sup>16</sup>. *TIF51A* (1.0 kb 5' flanking region, 0.5 kb ORF, and 0.5 kb 3' flanking region) was amplified from genomic DNA using primers P3 (5'-CTAGTCTAGAGCTGTCTCTACTAACAGATCTT GGCGTTTTGAATCGTG-3') and P4 (5'-CTAGTCTAGAGAGAAAAATAAC GACAACTGCAGGACTCGAACCTGCGCG-3'). The resultant PCR product was subcloned into XbaI-digested single-copy-number *LEU2* (YCplac111) and *URA3* (YCplac33) vectors or high-copy-number *LEU2* vector YEplac181 (ref. 31) generating plasmids pC3287, pC3288 and pC3289, respectively. Plasmid pC3290 was generated by inserting a Flag epitope tag at the C terminus of the *TIF51A* ORF in pC3289 using primers P10 (5'-AGGAAGCTGCTAGAACCGATGATT ATAAAGATGATGATGATAAATAAACCGGTT-3') and P11 (5'-ATCGGTTC TAGCAGCTTCCTTGAAGGAGATGG-3') and the GeneTailor Site-Directed Mutagenesis System (Invitrogen). Likewise, a K51R mutant derivative of pC3290 was constructed by site-directed mutagenesis using primers P14 (5'-CA TGTCACCTCTAAGACTGGTAGGCACGGTCACGC-3') and P15 (5'-ACCA GTCTTAGAAGTGACATGTCGACAATCTTAC-3'), generating the plasmid pC3291.

To randomly mutate the *TIF51A* ORF, we first used site-directed mutagenesis and the primers P16 (5'-ACACACACAAATACCAACTCTCGAGACAATGT CTGACG-3') and P17 (5'-GAGTTGGTATTTGTGTGTTGGGAGTCTA TG-3') to create an XhoI site immediately upstream of the *TIF51A* ORF and primers P18 (5'-CGATTAAACCGGTTAATCATCAGCGCCGGGATATAAA TG-3') and P19 (5'-TGATGTTAACCGGTTAATCGGTTCTAGCAGCT-3') to create an EagI site immediately downstream of the *TIF51A* ORF in pC3287, generating plasmid pC3292. The *TIF51A* ORF in pC3287 was amplified by error-prone PCR using primers P12 (5'-GATCCGCTCGAGACAATGTCTGACG AAGAACATACCTTTG-3') and P13 (5'-GATCTCGGCCGTGATGTTAAC CGGTTAATCGGTTCTAGC-3'). The PCR products were digested with XhoI and EagI and cloned into the XhoI/EagI site of pC3292, generating a *tif51a* mutant library. The plasmid pC3294 containing the *tif51a*(S149P) mutant was generated by site-directed mutagenesis of the plasmid pC3292 using primers P20 (5'-CTATGGGTGAAGAAGCCGCCATCCCCTCAAGGAAGCT-3') and P21 (5'-GATGCGCGCTTCTTACCCATAGCGGAGATGAT-3').

Yeast strain J700 was derived from H2557 by linearizing plasmid pAJ2 (ref. 16) with PmeI and integrating it at the *UBR1* locus. Correct pAJ2 integration was verified by PCR amplification of the *UBR1* locus and also by demonstrating galactose-dependent overexpression of Myc-tagged *UBR1* in J700. Strain J700 was transformed with the single-copy-number *URA3* plasmid YCplac33 to generate strain J713. The *TIF51A* degen mutant (*tif51a*-*td*) strain J701 was generated by integrating pC3286 (linearized with SalI) at the *TIF51A* locus in J700. The *tif51a*-*td* allele in J701 was confirmed by PCR analysis of genomic DNA and by western blot analysis of WCEs from cells grown under permissive conditions (raffinose medium supplemented with copper in which the *CUP1* promoter is induced and *UBR1* expression is low) versus non-permissive conditions (galactose medium lacking copper in which the *CUP1* promoter is repressed and overexpression of *UBR1* triggers proteasomal degradation of the degen protein) using rabbit polyclonal antiserum against yeast eIF5A.

The *tif51b*::*KanMX4* module from the *tif51b* (YJR047c) knockout strain in the yeast genome deletion collection was amplified by PCR using primers P5 (5'-GTTACCCTGAATCATATTGACGATGTCGTCTCACACGGA-3') and P6 (5'-CACCCCTCGTCGCAAAAAATTTTGACTTCTACTCTTT-3'), which hybridize ~600 bp upstream and downstream, respectively, of the *TIF51B* ORF. The PCR product was used to delete *TIF51B* in the haploid strains H2557, J701 and J713, generating strains J691, J702 and J714, respectively. The *KanMX4* marker in J691 was replaced by a *NAT* (nourseothricin-resistance) marker by transformation with EcoRI-digested p4399 (ref. 32) generating strain J692 (*tif51b*::*NAT*).

The haploid J692 strain was converted into a diploid strain (J693) by introducing a low-copy-number *URA3* plasmid carrying the *HO* gene. After isolation of a diploid transformant, cells lacking the *HO* plasmid were isolated on medium containing 5-fluoroorotic acid (5-FOA). Subsequently, the diploid strain J693 was transformed with the low-copy-number *URA3*, *TIF51A* plasmid pC3288, generating strain J694. The *tif51a*::*KanMX4* cassette from the heterozygous *tif51a*Δ strain from the yeast genome deletion collection (Open Biosystems) was amplified by PCR using primers P7 (5'-CGGGTACCGTGTATCCGAA GAGTCACTACCAAAAAC-3') and P8 (5'-GCATACATGCATGCTCCTTTT CAAGATCATCACCGTCGTCATCAC-3'), which hybridize ~1.5 kb upstream

and 1 kb downstream, respectively, of the *TIF51A* ORF. This PCR product was used to replace the *TIF51A* gene in J694 with a *tif51a*::*KanMX* cassette, creating strain J695. Strain J695 was sporulated, and subjected to tetrad analysis. Two of the four haploid spores carried the *tif51a*::*KanMX* allele in the chromosome and were unable to lose the *URA3*, *TIF51A* plasmid (unable to grow on medium containing 5-FOA). One of the two haploid strains, J696 (*tif51b*::*NAT* *tif51a*::*KANMX4* pC3288[*TIF51A*, *URA3*]), was transformed with the randomly mutated *tif51a* ORF library in pC3292 (low-copy-number *LEU2*). Individual transformants were patched on synthetic dextrose (SD) medium supplemented with essential nutrients, and then replica-printed to two 5-FOA plates, which were incubated at 25 °C and 37 °C, respectively. The *tif51a* mutant plasmids were isolated from cells that grew well at 25 °C but failed to grow at 37 °C. The isolated plasmids were re-tested to confirm that the mutant phenotype was plasmid-associated, and then sequenced to identify the mutation(s). One of the best temperature-sensitive mutants, *tif51a*(D63V), was selected for further analysis.

Strains J697, J698 and J699 were obtained by plasmid-shuffling after transformation of strain J696 with the *TIF51A* (pC3287), *tif51a*(D63V) (pC3293) and *tif51a*(S149P) (pC3294) plasmids, respectively. To purify Flag-tagged eIF5A (pC3290) and eIF5A(K51R) (pC3291), the indicated plasmids were introduced into strain J702, generating strains J703 and J704, respectively.

**In vivo [<sup>35</sup>S]Met incorporation.** Liquid cultures of yeast strains were grown under permissive conditions (SC medium lacking methionine and containing 2% raffinose and 100 μM CuSO<sub>4</sub>) at 25 °C to an A<sub>600</sub> of ~1.0. Cells were washed at room temperature (25 °C) to remove Cu<sup>2+</sup>, transferred to nonpermissive conditions (SC medium lacking methionine and containing 2% galactose and lacking CuSO<sub>4</sub>), and incubated for 14 h at 25 °C to A<sub>600</sub> of 0.6–1.0. To start the labelling, 50 μM unlabelled methionine and 1 μCi ml<sup>-1</sup> [<sup>35</sup>S]Met was added to each culture. At 15-min intervals, the optical density of each culture was determined, 1-ml aliquots were removed, and [<sup>35</sup>S]Met incorporation was monitored by trichloroacetic acid (TCA) precipitation.

**Polysome analysis.** Liquid cultures of cells were either treated with 50 μg ml<sup>-1</sup> CHX for 5 min before collection or left untreated, transferred to a 500-ml centrifuge bottle containing shaved ice, pelleted and washed with 10 ml of Buffer B (20 mM Tris, pH 7.5, 50 mM KCl, 10 mM MgCl<sub>2</sub>) supplemented with EDTA-free protease inhibitor tablet and 1 mM DTT. For CHX-treated cells, the same concentration of CHX was present in all steps thereafter. Cell pellets were resuspended in an equal volume of Buffer B, and then an equal volume of glass beads was added, and cells were broken by vigorous mixing on a vortex. After clarification, five A<sub>260</sub> units of WCEs were layered on 4.5–45% sucrose gradients prepared in 20 mM Tris, pH 7.5, 50 mM KCl, 10 mM MgCl<sub>2</sub> and 1 mM DTT, and then centrifuged for 2.5 h at 260,000g in a Beckman SW41 rotor. Gradients were fractionated while monitoring absorbance at A<sub>254</sub>.

**Ribosome transit time measurement.** Assays were performed as described previously<sup>21,33</sup> with modifications as detailed below. Yeast cells grown at 25 °C in SC-Met medium were shifted to prewarmed (36 °C) medium, and, after 2 h, 0.1 μCi ml<sup>-1</sup> [<sup>35</sup>S]Met was added. At regular intervals, 10-ml aliquots were removed and mixed with 200 μg ml<sup>-1</sup> CHX. Cells were resuspended in 0.2 ml cold lysis buffer (20 mM HEPES, pH 7.4, 2 mM magnesium acetate, 0.1 M potassium acetate, 0.5 mM DTT, 0.1 mg ml<sup>-1</sup> CHX) and broken with glass beads by vigorous mixing on a vortex for 15 min. Cell lysates and two washes of the glass beads were combined and clarified by centrifugation. Total (nascent plus completed) protein synthesis was determined by combining 0.5 ml lysate, 0.4 ml 60% sucrose and 0.5 ml 50% TCA. Mixtures were incubated for 10 min at 90 °C, 10 min on ice, and then collected on a GF/C filter that was pre-washed with 2 mM unlabelled Met. After washes with 10% TCA and then acetone, filters were dried and counted by liquid scintillation. Completed protein synthesis was determined by layering 0.5 ml lysate on a 0.4 ml 60% sucrose cushion and by pelleting ribosomes by centrifugation at 132,000g for 30 min in a Beckman TLA120.2 rotor. Radioactivity in the supernatant fraction (completed proteins) was determined as described previously.

**Purification of eIF5A and eIF5A(K51R).** C-terminal Flag-tagged proteins were overexpressed in yeast strains J703 and J704. After growth to mid-log phase under permissive conditions, cells were washed and then incubated under non-permissive conditions for 6 h to A<sub>600</sub> of ~3.0. Cells were washed, resuspended in Flag binding buffer (20 mM sodium-phosphate, pH 7, 500 mM NaCl, 0.1% Triton X-100, 1 mM phenylmethylsulphonyl fluoride and EDTA-free complete protease inhibitor tablet) and broken by high-speed mixing with glass beads. WCEs were mixed overnight with anti-Flag-M2 affinity gel (50% slurry, Sigma). After washing, bound proteins were eluted with Flag binding buffer containing 400 μg ml<sup>-1</sup> Flag-peptide, and dialysed against 20 mM HEPES, pH 7.5, 10% glycerol, 150 mM KCl and 2 mM DTT.

**Yeast in vitro translation assays.** *In vitro* assays of luciferase mRNA translation in crude extracts from yeast cells were performed as previously described<sup>34</sup> using



capped and polyadenylated luciferase mRNA transcribed *in vitro* from plasmid T7 LUC<sub>50</sub> (50-base poly(A) tail)<sup>35</sup>.

**Reconstituted translation elongation assays.** 80S initiation complexes were assembled essentially as described<sup>22</sup> with the following modifications. A different reaction buffer (20 mM Tris-Cl, pH 7.5, 100 mM potassium acetate, pH 7.6, 2.5 mM magnesium acetate, 0.25 mM spermidine, 2 mM DTT) was used and tRNA<sub>i</sub><sup>Met</sup> was labelled exclusively with [<sup>35</sup>S]Met. In addition, the mRNA contained the ORF sequence AUGUUCUUCUAA. After assembly, initiation complexes were layered onto reaction buffer containing 1.1 M sucrose and centrifuged for 1 h at 424,000g in a TLA-100.3 rotor. Complexes were resuspended in reaction buffer, flash-frozen, and stored at -80 °C. For each elongation reaction, Phe-tRNA<sup>Phe</sup> ternary complex was prepared. Each batch of ternary complex contained, as indicated in Fig. 3a, the following reagents: 45 pmol eEF1A, 21.6 pmol Phe-tRNA<sup>Phe</sup>, 11.25 pmol eEF2, 11.25 pmol eEF3, 20 pmol eIF5A (or eIF5A(K51R)), 2 mM GTP, 2 mM ATP and 1× reaction buffer, and was incubated at 26 °C for 15 min. At the end of this incubation, the ternary complex was mixed with a single aliquot of thawed initiation complex (~0.2 pmol), and incubated at 26 °C. Aliquots were removed at the indicated times and quenched in 0.25 M KOH. Reaction products were separated by electrophoresis on cellulose TLC plates (pyridine-acetate buffer, pH 2.8; 1,400 V, ~40 min). [<sup>35</sup>S]Met-containing reaction products were detected by phosphorimaging and quantified using ImageQuant 5.2 software (GE Healthcare Life Sciences).

**Reconstituted translation termination assays.** Initiation complexes were assembled as described previously, except that the mRNA coding sequence was AUGUUCUAA. Simultaneously, Phe-tRNA<sup>Phe</sup> ternary complex was assembled. Equal volumes of Phe-tRNA<sup>Phe</sup> ternary complex and initiation complex were mixed and incubated at 26 °C for 15 min. The elongation complexes were then pelleted and stored as described previously. Release factor complexes were assembled and incubated at 26 °C for 15 min, with the following components: 40 pmol eRF1, 40 pmol eRF3, 40 pmol eIF5A (or eIF5A(K51R)), 1.3 mM GTP and 1× reaction buffer. Single aliquots of elongation complex were mixed with release factor complex, and aliquots were withdrawn at the indicated times and quenched in 30% formic acid. Reaction products were separated, detected and quantified as described previously.

31. Gietz, R. D. & Sugino, A. New yeast-*Escherichia coli* shuttle vectors constructed with *in vitro* mutagenized yeast genes lacking six-base pair restriction sites. *Gene* **74**, 527–534 (1988).
32. Tong, A. H. *et al.* Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science* **294**, 2364–2368 (2001).
33. Shenton, D. *et al.* Global translational responses to oxidative stress impact upon multiple levels of protein synthesis. *J. Biol. Chem.* **281**, 29011–29021 (2006).
34. Tarun, S. Z. & Sachs, A. B. A common function for mRNA 5' and 3' ends in translation initiation in yeast. *Genes Dev.* **9**, 2997–3007 (1995).
35. Gallie, D. R., Feder, J. N., Schimke, R. T. & Walbot, V. Post-transcriptional regulation in higher eukaryotes: the role of the reporter gene in controlling expression. *Mol. Gen. Genet.* **228**, 258–264 (1991).

## CORRIGENDUM

doi:10.1038/nature08037

**Temperature-dependent thermal diffusivity  
of the Earth's crust and implications for  
magmatism**

Alan G. Whittington, Anne M. Hofmeister &amp; Peter I. Nabelek

*Nature* 458, 319–321 (2009)

In equation (3) of this Letter,  $10^{-6}$  should be  $10^6$ . The corrected equation is:  $C_{p,\text{crust}}(T < 846 \text{ K}) = 199.50 + 0.0857T - 5.0 \times 10^6 T^{-2}$ .

## CORRIGENDUM

doi:10.1038/nature08038

**New role of bone morphogenetic protein 7 in brown adipogenesis and energy expenditure**

Yu-Hua Tseng, Efi Kokkotou, Tim J. Schulz, Tian Lian Huang,  
Jonathon N. Winnay, Cullen M. Taniguchi, Thien T. Tran, Ryo Suzuki,  
Daniel O. Espinoza, Yuji Yamamoto, Molly J. Ahrens,  
Andrew T. Dudley, Andrew W. Norris, Rohit N. Kulkarni  
& C. Ronald Kahn

*Nature* 454, 1000–1004 (2008)

---

In the Methods Summary of this Letter, the concentrations of IBMX and dexamethasone were incorrectly listed as 0.5  $\mu$ M and 5 mM, respectively. The correct concentrations are 0.5 mM IBMX and 5  $\mu$ M dexamethasone.



# GOING NUCLEAR

Workforce shortages could slow the growth of an industry poised for a comeback. **Quirin Schiermeier** reports.

**F**rançois Perchet, an electrical engineer by training, has seen the ebb and flow of the nuclear power industry in the course of his long career. He spent more than 30 years with the French electricity company EDF, which operates the country's 58 nuclear reactors, and knew most of the 19 nuclear plant sites in France. He worked in operations management, then in maintenance and repair, and later in probability-based safety studies. But last December, he crossed the English Channel for a two-year assignment as programme coordinator with the World Nuclear University (WNU) in London, where he helps promote nuclear training and education throughout the world.

Now Perchet informs the public about the risks and benefits of nuclear power, an industry with big image problems. The stigma of nuclear power as a dangerous and toxic ticking time bomb has not only stymied industry investments and government funding but, critically, dissuaded young physicists, engineers and technicians from entering the field — so much so that the shortage threatens to thwart a potential nuclear renaissance.

But Perchet believes that attitudes towards the industry and its associated professions are changing. "In these difficult times, people start thinking about jobs for life," he says. "The nuclear power industry can't offer the huge salaries and bonuses that some bankers have been paid. But we can offer safe positions and a broad range of interesting career opportunities, from management to reactor development."

Like Perchet, most nuclear engineers

worldwide were recruited in the 1970s, before the near meltdown in 1979 at Three Mile Island in Pennsylvania. The Chernobyl disaster in the Soviet Union seven years later destroyed any remaining public faith in the safety of nuclear power and shattered the reputation of the nuclear industry as a whole. Nevertheless, there are 439 commercial nuclear-power reactors in 31 countries currently providing some 15% of total electricity supply worldwide, and the industry seems to have survived the fallout of these debacles as well as more minor accidents.

Still, the decades-long recruitment drought means that reactor manufacturers and operators are now facing the challenge of replacing almost half of their ageing workforce, including tens of thousands of qualified scientists and engineers around the world who are set to retire in the next 10 years.

Nuclear power is a major ingredient in the energy mix of many industrialized nations, most notably in France, where 77% of electricity comes from nuclear plants. In the United States, 104 reactors operated by 32 licensed utilities generate around one-fifth of the nation's electricity. Ukraine (48%), Sweden (46%), South Korea (35%), Japan (28%), Germany (22%), Russia (16%) and the United Kingdom (15%) all depend substantially on nuclear power as

well. Growth, however, has been mainly confined to Asia of late; no new plants have been built in the United States or western Europe for more than 25 years.

Despite its image problems, many countries have continued conducting nuclear science over the years, including research into reactor design and nuclear safety. In 2001, 13 nations formed the Generation IV International Forum to coordinate efforts for the development of a next-generation reactor.

Roughly 3,000 scientists worldwide are involved in generation IV reactor development; others are working in safety research.

Nuclear energy is now set for a comeback, with global warming from excessive fossil-fuel consumption accelerating and growing concerns over future energy security. Recently, many Western countries, including the United States, United Kingdom, France, Finland and Sweden, have abandoned *de facto* moratoria on building new plants. China and India also plan to extend their as yet small nuclear capacities. Plans exist to build at least 127 new reactors by 2020: China, India, Russia, Japan and the United

States, for example, plan more than ten new reactors each. Other interested countries include Brazil, South Africa, Vietnam, South Korea, Poland, Ukraine and Thailand. Under the most far-reaching scenario described in



**"There is a strong and increasingly competitive job market emerging."**  
— Carol Berrigan

H. KEHRER/ZEEFA/CORBIS

NUCLEAR ENERGY INSTITUTE

the report *Nuclear Energy Outlook*, published last year by the Organisation for Economic Cooperation and Development's Nuclear Energy Agency, up to 1,400 reactors of the size commonly in use today — producing 1,000–1,500 megawatts of electric power — would be in operation by 2050 worldwide.

The search is on for the scientists and engineers who will design, build and operate these nuclear reactors, but also for the experts who will guide uranium mining, design and partition nuclear fuels (that is, transmute radioactive waste into non-radioactive elements), dispose of radioactive waste, and protect nuclear workers and the world at large from harmful radiation. “There is a strong and increasingly competitive job market emerging,” says Carol Berrigan, who monitors the nuclear workforce for the Nuclear Energy Institute, an industry group based in Washington DC.

Academic institutions have responded to the nuclear comeback. Many universities have revived nuclear programmes that closed in the 1980s and 1990s because of lack of demand. In the United States alone, six new departments or programmes have been created in recent years at universities in Colorado, South Carolina, Texas and Virginia, and there are plans to set up programmes at the University of California, Los Angeles, and possibly also at the California Institute of Technology in Pasadena. Since 1999, enrolment in undergraduate nuclear engineering programmes in the United States has grown from 470 to around 2,000, while graduate enrolments have climbed from 220 to around 1,200, Berrigan found.

In Europe, advanced training and master's programmes in nuclear engineering exist, for example, at the University of Paris South 11, the Swiss Federal Institute of Technology in Zurich and Lausanne, and the Karlsruhe Institute of Technology (KIT) in Germany. In the United Kingdom, vocational training

and advanced courses in nuclear engineering are also offered by industry agencies such as Cogent and the National Skills Academy for Nuclear.

When it comes to exploiting that education, prospects in the US job market are promising. Graduate and postgraduate students interested in pursuing a career in the nuclear industry have ample choices between jobs at government labs, with private plant operators, in reactor construction or in the defence business, says Michael Corradini, head of nuclear engineering at the University of Wisconsin–Madison. “It's an incredibly healthy and vibrant environment for us right now, despite the economic downturn,” he says. “Most of my students can choose from at least a handful of job offerings; ten years ago they were glad when they got one.” Students often start with bachelor's degrees in civil, chemical or process engineering, then specialize in nuclear engineering as part of a master's degree.

Employment prospects in many regions of Europe are just as encouraging. “If a talented student turns up today, he or she will have a work contract tomorrow, either from me or from a company we're collaborating with,” says Joachim Knebel, head of nuclear safety research and acting head of neutron physics and reactor technology at KIT. The Areva Nuclear Professional School was established at KIT in February, funded by Areva, the Paris-based company that is the world's largest manufacturer of reactor components and nuclear fuel. It provides postgraduate training for young scientists and engineers specializing in various fields of nuclear

engineering. The 30 PhD students enrolled at the school at any one time are paid by Areva and have a guarantee that they will be employed when their training is completed. Areva, with a workforce of 75,000 and €13.2

billion (US\$17.2 billion) in global sales, is the market leader in the field and is now hiring about 1,000 people a month globally. Engineers of all sorts, from civil, electromechanical, chemical and process engineers to automation and computation experts, account for roughly one-half of newly hired staff.

Ongoing technological challenges should provide work for the most enterprising. Generation IV nuclear reactors are intended to replace current types such as the European Pressurized Reactor (also called Evolutionary Power Reactor or just EPR), a generation III reactor design that uses water under high pressure as coolant.

A mammoth research and development challenge ahead is the next generation of thermal and fast reactors (where the chain reaction is sustained by fast neutrons), which are designed to produce significantly less radioactive waste, to cost less and to be less vulnerable to violent attack. In principle, these systems would partition and reprocess nuclear waste repeatedly, which will require new technologies and new materials. These reactors could become commercially available in 20 to 30 years' time.

With the problem of safe nuclear-waste disposal unresolved, and the no less worrisome risk of proliferation of weapon-grade uranium and plutonium, nuclear power has not ceased to be controversial. But resistance to the peaceful use of nuclear energy, commonplace in global youth counterculture from the 1960s to the 1980s, has lost its fury. “Today's generation of engineering students couldn't care less about the nuclear resentments of old,” says Knebel.

Significant advances in reactor safety and nuclear fuel reprocessing technologies have helped. But it's the risk of unabated global warming that has made carbon-free nuclear power more acceptable, notes Annette Heinzel, a 35-year-old materials scientist at KIT who develops lead-cooled steel for the next reactor generation. “Back in 1999, most of my friends were a bit shocked when I left geology to pursue a career in nuclear reactor technology,” she says. “Ten years on, things have changed quite a lot. People have become more open-minded and most find it easy now to accept what I'm doing.”

**Quirin Schiermeier is Nature's Germany correspondent.**



**“Today's students couldn't care less about the nuclear resentments of old.”**  
— Joachim Knebel



The Penly nuclear power plant, located northeast of Dieppe, France.



# Ice blue

Escape from your cells.

**Paula R. Stiles**

Doing the Antarctic cryosphere slides is addictive. I take one out and click it into my headset. It's from Lake Vostok, two miles under the ice at the bottom of the world. Instantly, I'm in blue darkness. I'm a tiny single-celled organism, made up of protoplasm and primitive cellular structures — nucleus, mitochondria, cytoplasm, cell wall. I extend the latter into a tentacle to explore turgid water and flakes of ice that feel like icebergs to my microscopic size. I give off a small glow that illuminates my surroundings. I see some other glows far off. Not sure if they're friend or foe, but I'm hungry, so I go after them.

"Jethro ... hey, Jethro, wake up."

Blinking automatically clicks me out of the slide. The slide ejects itself, sticking out of the side of my headset. I remove it and lay it on the table next to me. I look up at my labmate, Sandra, still blinking. My eyes feel dry.

"What?" I say. It takes time to gear up from a single-celled organism to one with too many cells to count.

"You all right?" Sandra asks.

"Sure, fine." I sit up and start typing my observations into my computer. You can plug a slide into a machine and get all the data, but the human brain is still the best way to filter that experience. One good thing about living on the Moon is that it makes it easier to feel as if you're under water. One-sixth of Earth gravity is like that. I undo my headset and gently push myself to my feet.

"You've been acting funny ever since those new slides came in from Antarctica. I have to tell you, Jethro, I'm a little worried."

"I don't see why. I'm just doing my job."

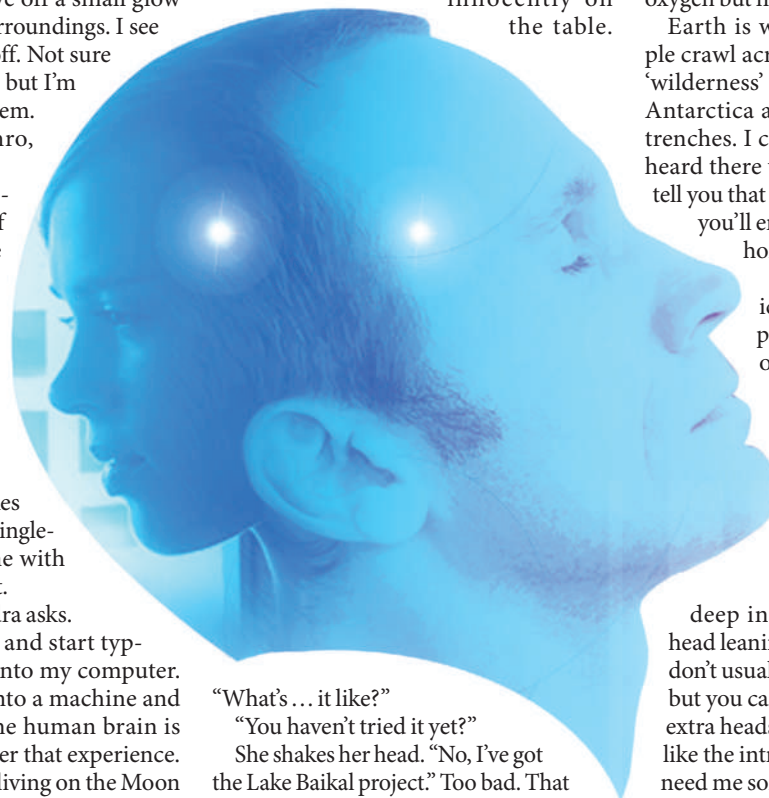
She shakes her head. As I look around, I try to see blues in the colour scheme of the lab. But everything is orange and pink, the better to keep us happy and energized. "It's more than a job for you," she says. "You're getting lost in there."

I shrug that off. "Would you rather I did an oxygen high or went outside without a spacesuit? Seems to me that doing my job

and enjoying it isn't the worst thing out there."

"Yes, but you're really hogging the slides." Now that I'm getting a closer look at her, she doesn't seem so much obsessed with the rules as something else. In fact, she looks downright pinched, like an oxygen addict. Oxygen is rationed here on the Moon. You have to pay for it and boy, is it expensive. The poor (the government likes to call them "more frugal") are chronically anoxic.

She glances towards the slide lying innocently on the table.



"What's ... it like?"

"You haven't tried it yet?"

She shakes her head. "No, I've got the Lake Baikal project." Too bad. That water is overloaded with ugly organisms and pollution — not much fun. "So, what's it like?"

"Blue. Quiet."

Her eyes glaze over. Oh, my, Sandra. Are you a solitude junkie, too? "That sounds ... nice."

"I'm not complaining." This is my cue to leave. Sandra's staring at that slide like a starving zoo lion at meat, but she's too shy to get acquainted with it until I'm safely gone.

"Anyway, I gotta get home. There's a new holo on tonight and I'm not sure my recorder's working..." This is a total lie. No holo out there can replace that cool blue. Still, share and share alike.

Sandra's attention is no longer on me; it's

on that slide. So much for worrying about my welfare. She's already sitting down and plugging it in as I slip out of the room.

The main office is a honeycomb of cubicles inhabited by plugged-in workers who look like termite grubs — all oblivious to me as I get into the elevator and head for the surface living quarters in the city. Moon buildings extend down, not up. The good thing about working in an office is that they circulate the air more regularly, to keep our efficiency up. It smells sharp and citrusy like air-conditioning chemicals, oxygen but not real air.

Earth is worse. Twenty billion people crawl across its face. Any remaining 'wilderness' hides underneath the ice of Antarctica and in the bottom of ocean trenches. I came to the Moon because I heard there was open space. They don't tell you that most of it isn't habitable, that you'll end up living in hives. Just like home.

I try to keep a bubble of ice-blue water in my head as I pass out into the main conduit of the city, a highway of people shoving, bumping, passing over each other. And the smell of stale air ... every belch, fart, cast-off skin flake, perfume, deodorant and hair oil.

I last about five minutes.

Back in the lab, Sandra is deep in Lake Vostok, eyes closed, head leaning back against her chair. We don't usually do it, except for team jobs, but you can 'share' a chip. I pull out the extra headset and plug in. Sandra won't like the intrusion, but she'll get over it. I need me some ice.

The water in the lake isn't water, exactly — more like fluid ice, warmed by the weight of two miles of continental ice sheet and geothermal heat from Earth's mantle. I oscillate to increase my bioluminescence. In the distance, I sense another organism, its pseudopod probing the icy lake ceiling. It senses me as well. Pausing, the organism waves its pseudopod at me in a human gesture no one-celled organism would make — Sandra. I wave back.

We go in opposite directions in the dark.

Paula Stiles has sold SF, fantasy and horror stories to *Strange Horizons* and *Writers of the Future* among others. Her online home is [www.geocities.com/rpcv.geo/other.html](http://www.geocities.com/rpcv.geo/other.html).

JACEY